PREDICTING THE ANTIGENIC EVOLUTION OF INFLUENZA VIRUSES WITH APPLICATION TO VACCINATION STRATEGY

DAVID JOSEPH PATTINSON

Department of Zoology
School of Biological Sciences
University of Cambridge

Jesus College

September 2019

This dissertation is submitted for the degree of Doctor of Philosophy.

SUPERVISOR
Derek J. Smith

LOCATION
Centre for Pathogen Evolution
Department of Zoology
University of Cambridge
Downing St.
Cambridge
CB2 3EJ, U.K.

TIME FRAME
September 2019

DECLARATION

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text.

It does not exceed the prescribed word limit for the relevant Degree Committee.

Cambridge, U.K., September 2019

David J. Pattinson

SUMMARY

Seasonal influenza viruses cause substantial worldwide mortality and morbidity every year. The evolution of these viruses comprise unique systems for studying natural evolutionary processes in real-time. Host immune systems recognise pathogens based on binding affinities between host antibodies and pathogen antigens. Pathogens with similar antibody binding are said to be *antigenically* similar. Seasonal influenza viruses evolve antigenically over a timescale observable by humans. Vertebrate immune responses also adapt rapidly, such that a single infection usually leaves a host protected for life against antigenically similar strains. In humans this imposes natural selection for antigenic novelty in wild virus populations.

Influenza vaccines contain virus antigen which elicit the production of antibodies that protect against antigenically similar strains. After major antigenic evolution in wild viruses, vaccines must be updated to remain effective. I use a simple model to show that even in best case scenarios current influenza vaccination strategies cannot completely avoid antigenic mismatch. I then present the first study that quantifies the link between degree of antigenic mismatch and vaccine effectiveness.

Knowledge of the molecular variation responsible for major antigenic change in natural influenza viruses has improved greatly in the last decade. I review this work, and present the application of a computational approach from the field of quantitative genetics to this problem. Subsequently, I show that patterns in biophysical features of substitutions responsible for major antigenic change are far from random, begging the question: how predictable are these amino acid substitutions? I answer this question by developing a computational framework to rank candidate substitutions by their biophysical similarity to substitutions responsible for major antigenic change and show that predictions can be made that are substantially better than chance selections.

Finally, I discuss the application of these rankings to advanced influenza vaccination strategies based on the principle of immunity management. I expand on how this work should be the basis for further investigation into the mechanisms that govern different components of influenza virus fitness and ultimately the antigenic evolution of seasonal influenza viruses.

To Mum and Dad.

ACKNOWLEDGMENTS

I feel very lucky to have spent the last five years doing work I love with colleagues who have become close friends. There are many people I would like to thank.

Derek, your dedication and ruthless approach to science is infectious. Thank you for including me in the research environment you have nurtured, and for giving me the time and space to finish this work. A more bizarre moment with you was whilst I was dressed in Barbara's tights, a blue bin liner and a cardboard mask as a bird of paradise at the EMC Christmas party. You sought me out in the middle of the crowded dance floor to talk about an idea you'd had for the rankings. I think we had seen the first bumps plots earlier that week. You engaged me in a full on scientific discussion in the midst of a mêlée of drunk Dutch jungle animals with the same seriousness and focus as if we were sat in S8. Please never change.

Terry, thank you for being so generous with your time, knowledge and wonderful idiocy. An especially big thank for 'steering the ship' in the last couple of years. Barbara and Sam, congratulations on finishing your degrees. Thank you for going through it with me, putting up with tea bags on the sink and making graphs together. To everyone else in my group, thanks for sharing your insights, criticisms, praise and for helping me so much.

Mum and Dad, thank you for teaching me to see the world the way I do. Thank you for the opportunities and love and support. Emma and Katie, thank you for the distractions and silliness and always being there.

Finally, thank you Hannah. You have been a rock at every stage of this. You pulled me through a long period of depression when finishing this thesis was unthinkable. You've been with me through long periods of slow progress. You've read every word in this thesis with care and always gave me a fresh perspective. You are entirely responsible for everything that stems from Chapter 2. Thank you for your time and interest and putting up with my constant stream of ill-posed questions. All the while you were finishing your own Ph.D. and setting up your own research group. Let's keep collaborating in work and life.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

AAP Amino Acid Polymorphism

AIC Akaike Information Criterion

BIC Bayesian Information Criterion

BLUP Best Linear Unbiased Predictor

CDC Centers for Disease Control and Prevention

cRNA Copy RNA

DNA Deoxyribonucleic Acid

FRA Focus Reduction Assay

GISRS Global Influenza Surveillance and Response System

GISAID Global Initiative on Sharing All Influenza Data

GWAS Genome Wide Association Study

HA Haemagglutinin

HI Haemagglutination Inhibition

IIV Inactivated Influenza Vaccine

ILI Influenza-Like Illness

KDE Kernel Density Estimation

LAIV Live Attenuated Influenza Vaccine

LMM Linear Mixed Model

MDCK Madin-Darby Canine Kidney

mRNA Messenger RNA

mvLMM Multivariate Linear Mixed Model

M1 Matrix protein 1

M2 Matrix protein 2

NA Neuraminidase

NEP Nuclear Export Protein

NIC National Influenza Centre

NS1 Non-structural protein 1

NS2 Non-structural protein 2

NP Nucleoprotein

OLS Ordinary Least Squares

OR Odds Ratio

PA Polymerase Acid

PB1 Polymerase Basic 1

PB2 Polymerase Basic 2

PCR Polymerase Chain Reaction

PRNT Plaque Reduction Neutralisation Test

RCT Randomly Controlled Trial

RNA Ribonucleic Acid

RNP Ribonucleoprotein

RR Relative Risk

RRM Realised Relationship Matrix

RBS Receptor Binding Site

SNP Single Nucleotide Polymorphism

TIV Trivalent Inactivated Vaccine

TND Test Negative Design

VE Vaccine Effectiveness

VIDRL Victorian Infectious Diseases Reference Laboratory

vRNA Viral RNA

WHO World Health Organisation

WLS Weighted Least Squares

AMINO ACID ONE LETTER CODES

A Alanine

C Cysteine

D Aspartic acid

E Glutamic acid

F Phenylalanine

G Glycine

H Histidine

I Isoleucine

K Lysine

L Leucine

M Methionine

N Asparagine

P Proline

Q Glutamine

R Arginine

S Serine

T Threonine

v Valine

w Tryptophan

y Tyrosine

MATHEMATICAL NOTATION

y Regular, lower case: a scalar.

$\mathbf{y}$ Bold, lower case: a one dimensional column vector.

$\mathbf{Y}$ Bold, upper case: a matrix.

$\mathcal{N}(\mu, \sigma)$ Normal distribution with mean, μ, and standard deviation σ.

$\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ Multivariate normal distribution with mean, $\boldsymbol{\mu}$, and covariance, $\boldsymbol{\Sigma}$.

$\mathcal{MN}(\mathbf{M}, \mathbf{U}, \mathbf{V})$ Matrix normal distribution with mean, $\mathbf{M}$, row covariance, $\mathbf{U}$, and column covariance, $\mathbf{V}$.

CONVENTIONS

MUTATION The change of a nucleotide base in DNA or RNA.

SUBSTITUTION The change of an amino acid in a protein.

XNY The substitution of amino acid X to amino acid Y at position N.

ΔN The residue at position N is deleted.

1 INTRODUCTION

The evolutionary biology of human seasonal influenza viruses is a rich and fascinating field, in which basic research can have major translational impact on global public health. Although notorious for their ability to evolve, recent discoveries have demonstrated that the antigenic evolution of seasonal influenza viruses is constrained in several ways. In this thesis I demonstrate how these constraints may be exploited in order to predict future antigenic variants of seasonal influenza viruses, and discuss novel vaccination strategies which may ameliorate the issue of antigenically mismatched influenza vaccines.

CHAPTER OUTLINE I first introduce the basic ecology of influenza viruses; how spillover events cause pandemics in secondary hosts producing new subtypes that circulate as seasonal viruses. I explain steps in influenza virus replication and how they are mediated by viral proteins. This forms important knowledge for understanding processes in the virus life cycle that can be blocked by host antibodies. I describe the impact of seasonal and pandemic influenza viruses on global human health and explain the role of antigenic variability in this. Then, I detail how antigenic phenotypes are measured and analysed using antigenic cartography. I introduce influenza vaccine technology and outline how a better understanding of influenza virus evolution would enable improved vaccination strategies.

Finally, I outline how the research in this thesis aims to 1) understand more rigorously the relationship between antigenic mismatch and Vaccine Effectiveness (VE) and 2) test how predictable the antigenic evolution of seasonal influenza viruses are and what the implications are for vaccination strategy.

1.1 INFLUENZA VIRUS BIOLOGY

Influenza viruses are part of the Orthomyxovirus family (the Orthomyxoviridae) which are characterised by a negative sense, single stranded and segmented Ribonucleic Acid (RNA) genome. Negative sense means that the genome is complementary to that of Messenger RNA (mRNA), so that upon entry into the nucleus, Viral RNA (vRNA) is transcribed directly into mRNA, from which viral proteins are translated. There are four types of influenza viruses: A, B, C and D. Types A and B both have eight genome segments and cause the majority of disease burden in humans (Krammer et al. 2018). Types C and D

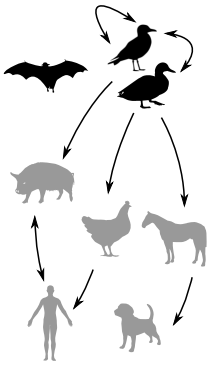
have seven genome segments and do not cause substantial disease in humans.

Within type A there are numerous subtypes that are categorised by the antigenic properties of the two surface glycoproteins: Haemagglutinin (HA) and Neuraminidase (NA). Subtypes are defined by their lack of cross reactivity to one another. For example, if a naive animal was infected with an A(H3N2) virus, its antisera would show very low or non-detectable reactivity to a virus from any other subtype.

1.1.1 *Nomenclature*

Subtype names concatenate the type of HA and NA, e.g., "H12N5". Names for specific virus strains were standardised in 1980 (World Health Organization 1980). For example "A/turkey/Ontario/7732/66 (H5N9)" refers to a type A, subtype H5N9, virus isolated from a turkey in Ontario, Canada in 1966. The generic ID for the virus is 7732. Human isolates omit the host field, e.g., "A/Hong Kong/1/68 (H3N2)".

1.1.2 *Primary reservoirs*



Primary (black) and secondary (grey) reservoirs for influenza A viruses.

The primary reservoir for influenza A viruses are aquatic birds in which many subtypes circulate continuously and cause little or no disease (Webster et al. 1992). Birds in the order Anseriformes (especially ducks, geese and swans) and Charadriiformes (especially gulls, terns and waders) are the main hosts (Fouchier and Guan 2013). To date, 16 different HA subtypes (H1-H16) and nine different NA subtypes (N1-N9) have been isolated from birds (Donatelli et al. 2016). The segmented genome facilitates gene reassortment whereby a host infected simultaneously with two strains can produce progeny virus that inherit segments from both infecting strains. This has given rise to the plethora of different combinations of HAs and NAs isolated from wild birds (Fouchier and Guan 2013).

Recently two novel subtypes were detected not in aquatic birds, but in new world bats (Tong et al. 2012; Tong et al. 2013). These bat viruses show deep phylogenetic divergence between each other and to other influenza A virus subtypes, suggesting that bats may also harbour additional subtypes (Tong et al. 2013).

1.1.3 *Secondary reservoirs*

Influenza viruses also infect other mammals including horses, whales, seals, cats, dogs, pigs, poultry and humans. All subtypes found in these secondary reservoirs are also found in aquatic birds. Viruses transmit between different hosts via cross-species transmission events. For example, phylogenetic analysis suggests that the A(H1N1) virus

that caused the 1918 influenza pandemic in humans originated from an avian source, and that after the pandemic it crossed from humans into swine (Worobey et al. 2014). Similar analyses have shown that the 2009 A(H1N1) influenza pandemic virus was transmitted from pigs to humans (Smith et al. 2009). Reassortment of the segmented genome often accompanies host switch events and the emergence of new strains. For example, a virus that is adapted to replicate efficiently in a mammalian host may acquire novel surface antigens from a different subtype, producing an antigenically novel, mammalian adapted strain. The emergence of the 1957 A(H2N2), 1968 A(H3N2) and 2009 A(H1N1) pandemic strains were associated with genetic reassortment events prior to their widespread circulation in humans (Lindstrom et al. 2004; Smith et al. 2009).

1.1.4 *Large scale evolutionary dynamics*

Influenza pandemics occur when a host shift event introduces an antigenically novel strain into a host reservoir. One reason that influenza pandemics can be so virulent is that the new host reservoir has very low or no standing immunity to the pandemic strain. Therefore, once infected the virus is able to replicate efficiently within a host without the host adaptive immune response interfering. On subsequent exposure to a similar strain, a host is able to mount an effective immune response and the infection is either unsuccessful or attenuated.

Components of a pathogen recognised by the immune system are called *antigens*, therefore 'similar' here specifically refers to *antigenic* similarity. Once a pandemic strain has infected most naive hosts in a population it can no longer replicate and spread as effectively. Therefore, host immune systems impose a natural selection for antigenic variation in the virus population. If antigenic variants evolve, hosts maintain a degree of cross protection from the prior exposure to the pandemic strain. Epidemics of seasonal influenza ensue whereby a particular antigenic variant circulates, immunity grows to it, and eventually subtly different antigenic variant emerges.

These dynamics can be summarised as reservoirs of subtypes in aquatic birds which spillover into other hosts causing pandemics; after the initial spread of the pandemic virus, seasonal dynamics begin whereby host immunity develops, selecting for continual antigenic evolution.

1.1.5 *Virus structure*

Influenza virus particles can be spherical with a diameter of approximately 120 nm, or filamentous, with a length of several micrometres (Chu et al. 1949; Choppin et al. 1960). Particles are contained by a lipid bilayer taken from the host cell during the budding process

Main sources on influenza virus structure and replication were Fields Virology (Fields et al. 2013) and the Textbook of Influenza (Webster et al. 2013).

(Figure 1.1). Embedded in the viral membrane are the HA and NA glycoproteins which coat the majority of the outside of the particle in a ratio of approximately three-to-one.

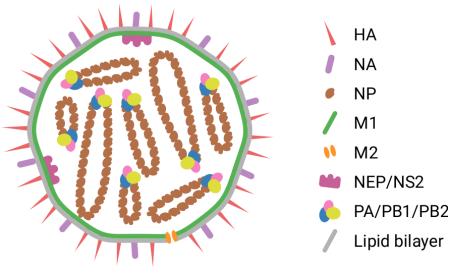


Figure 1.1: Schematic influenza virus particle. The virion is contained by a lipid bilayer taken from the parent cell during budding. The Haemagglutinin (HA) and Neuraminidase (NA) spike proteins are embedded within the viral membrane and project away from the virus. Matrix protein 2 (M_2) forms an ion channel across the membrane, allowing external pH changes to propagate inside the virus. Matrix protein 1 (M_1) forms a structural layer directly within the virus membrane. The Nuclear Export Protein (NEP)/Non-structural protein 2 (NS_2) complex is associated with the internal face of the membrane and the M_1 layer. The core of the virion contains eight Ribonucleoproteins (RNPs) that are each composed of single RNA genome segments coated with Nucleoprotein (NP) and bound at either end by the Polymerase Acid (PA), Polymerase Basic 1 (PB_1), Polymerase Basic 2 (PB_2) polymerase complex.

HA binds to sialic acids expressed on the surface of host cells (Wiley et al. 1981), and is responsible for fusing the viral and endosomal membranes during replication (Skehel and Wiley 2000). Nascent HA is a homotrimer; each monomer is expressed as a single polypeptide (HA0) which after viral budding is cleaved by host proteases into HA1 and HA2. HA1 comprises the membrane-distal globular head domain containing the Receptor Binding Site (RBS). Structurally, HA1 is composed of several β-sheets. HA2 comprises the stem and transmembrane domains and also houses the fusion peptide and fusion machinery. Structurally, it is composed of two main α-helices linked by the B-loop. The B-loop has a strong propensity for an α helix conformation itself, which acts as a molecular spring in the fusion mechanism.

β-sheets are common secondary structure motifs found in proteins. They consist of parallel peptide chains held together by regular sequences of hydrogen bonds between backbone NH and CO groups.

NA has the complementary role of cleaving sialic acid moieties (Air and Laver 1989). Recently the clear division of labour between HA and NA has been blurred by the discovery of A(H3N2) strains in which the NA possesses haemagglutination activity (Lin et al. 2010; Mögling et al. 2017). The M_2 protein channel is also embedded in the viral membrane and acts as a hydrogen channel. M_1 coats the inside of the viral membrane; the NEP/NS_2 complex is associated with the inside of the viral membrane and M_1. The eight vRNA segments are coated by NP which has a non-specific RNA binding groove and a net positive charge which facilitates binding to RNA, which has a negative charge.

The ends of the vRNA are associated with the PA/PB1/PB2 polymerase complex which pulls the vRNA into a distinctive loop. A single vRNA sheathed in NP and bound to the polymerase complex is referred to as an RNP.

1.1.6 *Replication*

Virus particles attach to host cells via interactions between HA and sialic acids expressed on the cell surface (Figure 1.2). Binding interactions between a single HA and sialic acids are relatively weak (Sauter et al. 1989; Sauter et al. 1992). Binding events between many HAs on a single virion with multiple sialic acids help initiate the engulfing process. Four different modes of entry have been identified for influenza viruses: via clathrin-coated pits, caveolae, macropinocytosis and non-clathrin, non-caveolae pathways (Fields et al. 2013). Independent of the specific mode of entry, the virus particle ends up within an endosome in the cell cytosol.

The endosome has a lower pH than the inside of the virus, so protons begin to cross the viral membrane through the M2 proton channel. The lower pH within the virus particle induces a conformational change in HA2, whereby two α helices that are folded back on each other and linked by the B-loop, form a single long α helix (Skehel and Wiley 2000; Colman and Lawrence 2003). Specifically, residue 106 of HA2, and surrounding residues in an inter-subunit cavity, become protonated which unclamps the B-loop allowing it to adopt its α helical confirmation (Xu and Wilson 2011). This conformational change moves the fusion peptide, which contains predominantly hydrophobic residues, outside of its pocket and into proximity with the endosomal membrane. The fusion peptide embeds in the endosomal membrane which facilitates the fusing of the viral and endosomal membranes. Eventually a pore opens and the RNPs are released into the cytosol.

NP contains nuclear localisation signals which hijack cellular machinery into trafficking RNPs to the nucleus where they are actively imported. Inside the nucleus the polymerase complex copies the negative sense vRNA into positive sense mRNA and Copy RNA (cRNA). vRNA is packaged in viral genomes and contains coding regions as well as 3‘ and 5‘ untranslated regions. cRNA is a complementary copy of the entire vRNA which enables amplification of more vRNA. mRNA does not contain the untranslated regions, it is 5‘ capped and polyadenylated and gets exported from the nucleus and acts as a template from which host ribosomes translate viral proteins. Virus polymerases require host cellular RNA polymerase II to initiate RNA replication.

Once translated, viral membrane proteins (HA, NA and M2) get incorporated into the cellular membrane. NEP, Non-structural protein 1 (NS1) and M1 also begin to accumulate in dense spots on the inside of the cellular membrane, and small pits begin to form. Untranslated

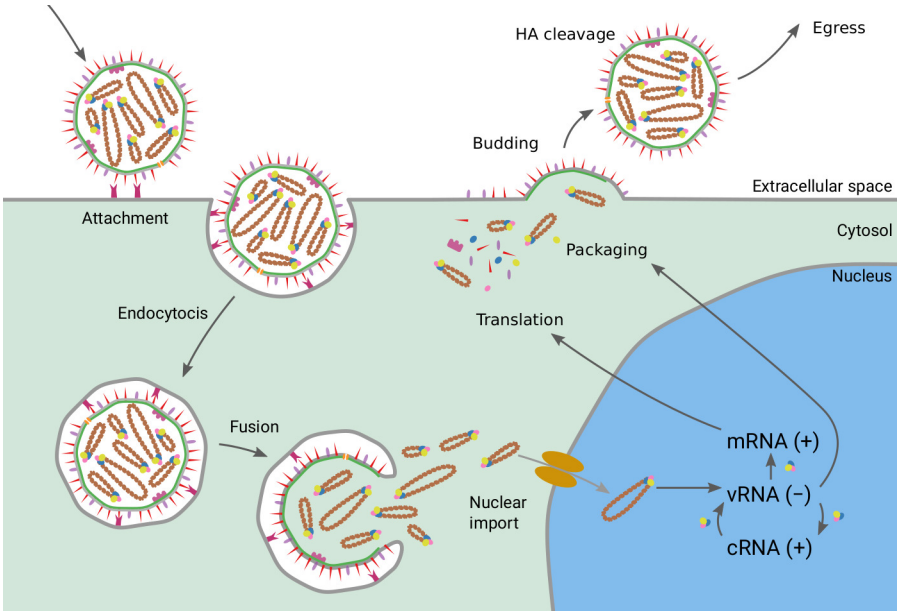


Figure 1.2: Influenza virus replication. Virions attach via interactions between HA and sialic acids expressed on the surface of host cells. Endocytosis causes the virion to become fully encapsulated within the host cell. Lower pH within the endosome causes a conformational change within HA which forces the viral and endosomal membranes to fuse and releases RNPs into the cytosol. RNPs are actively transported to, and imported into, the nucleus, where positive sense cRNA and mRNA are transcribed. Additional copies of vRNA are generated from the cRNA and exported to the cytosol. Viral mRNA is also exported to the cytosol, from which viral proteins are translated. Viral membrane proteins are incorporated into the host membrane. Denser regions of virus proteins in the membrane form pits, which eventually bud into free virions. Host proteases cleave HA0 to produce mature HA1 and HA2 which activates the HA fusion sensitivity to low pH.

regions of vRNA contain packaging signals which identify segments and allow complete sets of segments to be incorporated into budding virions (White et al. 2019). Host proteases in the extracellular space cleave the HA0 precursor into HA1 and HA2 which activates the pH-sensitive fusion mechanism. This prevents premature release of the fusion peptide before virions are fully budded. Sialic acid-destroying activity of the NA cleaves external sialic acids, enabling efficient release of virus particles.

1.2 HUMAN INFLUENZA VIRUSES

In humans, influenza viruses cause seasonal epidemics and occasional global pandemics. Influenza cases are variable, but symptoms can include fever, cough, sore throat, muscle ache, runny nose, fatigue, headaches, and occasionally vomiting and diarrhoea. Collectively these

symptoms are referred to as Influenza-Like Illness (ILI). The severity of cases range from being asymptomatic to fatal. Asymptomatic infections can be inferred by detecting new antibodies or by detecting vRNA in samples. The virus is spread by direct contact, indirect contact via contaminated surfaces, and airborne transmission via aerosols or larger respiratory droplets (Schrauwen et al. 2014; Neumann and Kawaoka 2015).

Only influenza types A, B and C circulate in humans (Krammer et al. 2018). The majority of the disease burden in humans is caused by types A and B, although occasionally type C causes ILI in children (Matsuzaki et al. 2006). Four seasonal subtypes currently circulate in the human population: A(H3N2), introduced by the 1968 pandemic; A(H1N1), introduced by the 2009 pandemic and two lineages of influenza B viruses. Human influenza B viruses split into two phylogenetically and antigenically distinct lineages in the late 1980s which are referred to as the B(Victoria) and B(Yamagata) lineages (Rota et al. 1992; Kanegae et al. 1990; McCullers et al. 2004). Unlike influenza A viruses, which infect many hosts, influenza B viruses are primarily found in humans, where they exist at lower prevalence, and tend to cause less severe disease than human type A viruses (Webster et al. 1992).

1.2.1 *Impact of influenza*

The devastating impact of past influenza pandemics on global public health are well-known. The last four influenza A pandemics killed on average 0.70 % of the global population (Table 1.1). These pandemics occurred over the course of a century, or one every 25 years, meaning that this value corresponds to approximately $0.70/25 = 0.028\%$ of the global population per year. In terms of attack rates, it is estimated that over half of school children in Hong Kong were infected during the first wave of the 2009 A(H1N1) pandemic (Wu et al. 2010) and that 28 % of the population of the U.S.A. was infected during the 1918 A(H1N1) pandemic (Frost 1920).

This calculation should be interpreted as a crude estimate. It uses only four observations which range in value across three orders of magnitude (Table 1.1).

The global impact of seasonal influenza is less well appreciated. The World Health Organisation (WHO) estimates that 5 % to 10 % of adults and 20 % to 30 % of children are infected globally each year (World Health Organisation 2014). Recently it was estimated that between 1999 and 2015, seasonal influenza killed 290,000-650,000 people worldwide each year (Iuliano et al. 2018) which corresponds to 0.004 % to 0.0088 % of the global population. The upper bound of this range is only 3.2 times smaller than the global yearly death toll of pandemic influenza, estimated in the previous paragraph. The impacts of seasonal influenza are felt disproportionately in countries with less well-developed health systems. An estimated 99 % of deaths

Pandemic	Population (billions)	Mortality (millions)	Mortality (%)
1918 A/H1N1	1.8	50	2.7
1957 A/H2N2	2.9	1.1	0.038
1968 A/H3N2	3.6	1.0	0.028
2009 A/H1N1	6.9	0.36	0.0053

Table 1.1: Global mortality of influenza A pandemics. Mortality estimates are expressed as counts and as a percentage of the world population when the pandemic occurred. Mortality estimates were obtained from Viboud et al. (2005), Taubenberger and Morens (2006), Dawood et al. (2012), and Viboud et al. (2016). World population data were obtained from Roser et al. (2019). All values are shown to two significant figures.

of children under five years of age caused by seasonal influenza occur in developing countries (Nair et al. 2011).

Summary statistics for pandemics are highly skewed by the 1918 pandemic. Most deaths during the 1918 pandemic occurred in the second of three pandemic waves late in 1918 which left a dramatic mark on the trend of increasing life expectancy in the U.K. at the time (Figure 1.3). If the 1918 pandemic had the same mortality as the average of the other three pandemics, then the total impact of seasonal influenza on global mortality would be approximately 6.9 times greater than the total impact of pandemic influenza. These statistics clearly have wide confidence intervals, and should not be taken as highly precise. Instead, they serve to highlight that seasonal influenza has a similar impact to pandemic influenza with respect to global mortality.

The economic burden of seasonal influenza is also substantial. A study estimated that lost earnings due to sickness and mortality plus direct medical costs totalled 87 billion dollars in the U.S.A. in 2003 (Molinari et al. 2007). Improving the effectiveness of influenza vaccines has potential to directly mitigate these economic burdens. It has been estimated that a one percent improvement in Vaccine Effectiveness (VE) would result in savings of $400 million p.a. in the U.S.A. alone (Mosterín 2014).

VE is introduced in detail in Subsection 3.1.2.

1.3 ANTIGENIC VARIABILITY

In 1926, it was shown that ferrets previously exposed to influenza viruses were protected from subsequent infections with the same strain (Smith et al. 1933). However, in 1936 it was demonstrated in mice that exposure to one strain did not protect against all other influenza virus strains (Magill and Francis 1936), and that therefore

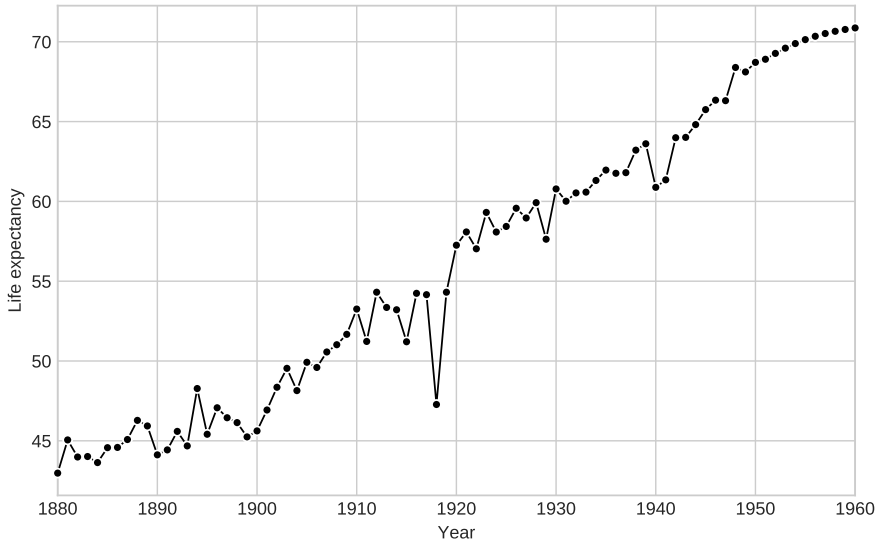


Figure 1.3: Life expectancy in the U.K. from 1880 to 1960. The drop in life expectancy in 1918 was caused by the 1918 influenza pandemic. Data from Roser (2019).

different antigenic variants must exist. *Homologous* and *heterologous* are terms that describe a subsequent exposure that either matches or differs from a prior exposure. Additional antigenic variants were identified in 1938 and homologous vaccination was shown to be more effective than heterologous vaccination (Smith and Andrewes 1938). Antigenic variation in natural influenza viruses meant that the first influenza vaccines had to be reformulated after only two years of use (Salk and Suriano 1949; Payne 1953).

The antigenic variability of influenza viruses is notably different to the many other pathogens which are antigenically static. There are many examples of antigenically static pathogens, including Variola virus (which causes small pox) and Measles virus. Effective vaccines have been developed for both of these viruses, and they only have to be administered once to confer life-long protection because the pathogens do not evolve antigenically. Variola virus was eradicated in a global vaccination effort in the mid-20th century, and vaccination caused global measles mortality to drop from 750,000 deaths in 2000 to 197,000 in 2007 (Philippe et al. 2009). Both of these achievements were possible due to the development of effective vaccines, and an antigenically static pathogen.

Recent localised increases in measles are associated with vaccine hesitancy which increases the risk of infection (Salmon et al. 1999; Salmon et al. 2015).

1.3.1 *Immunological assays*

Assays for measuring antigenic phenotypes have been in development since the antigenic variability of influenza viruses was first established. In 1941, George Hirst noticed that the allantoic fluid of chick embryos infected with influenza viruses agglutinated red blood cells

Embryos are associated with four membranes: the yolk sac, amnion, allantois and chorion. The allantois contains allantoic fluid.

(Hirst 1941). Antiserum is blood with all cells and clotting factors removed; it contains antibodies, antigens, hormones, and electrolytes. Hirst showed that the agglutination was blocked when antisera from a previously infected chick was added to the allantoic fluid. He recognised that this reaction could act as an endpoint to measure binding between an antisera and a particular influenza virus (Hirst 1942). Since then, the Haemagglutination Inhibition (HI) assay has remained largely unchanged for almost 80 years. Serial dilutions of antiserum are made, typically across a 96-well plate. Virus and red blood cells are added to the plate which is then incubated for a set time. The HI titre is the dilution at which the antiserum is no longer able to block the agglutination of the red blood cells. Hirst showed that HI titres correlated with mouse lethal infection titres, and neutralisation titres (Hirst 1942).

Many other assays are used to measure antigenic phenotypes. Virus neutralisation is the process by which antiserum blocks some aspect of virus function which reduces virus growth. Neutralisation assays measure the extent to which a virus is neutralised by a particular antiserum. A common neutralisation assay is the Plaque Reduction Neutralisation Test (PRNT), which is also known as the Focus Reduction Assay (FRA). Varying dilutions of antisera are mixed with virus and spread on cell monolayers; the concentration of antiserum that is able to reduce the plaque growth count by 50 % is referred to as the $PRNT_{50}$.

PRNTs which take several days are more time-consuming, cumbersome and expensive than HI assays, which take only several hours. In some respects, measuring an antiserum's ability to neutralise virus, rather than simply to disrupt HA binding, may be a biologically more useful measurement. Viruses may be neutralised by interfering with any critical step in virus replication, whereas the HI assay measures only the extent to which an antiserum blocks the binding of influenza viruses to red blood cells. Despite the relative crudeness of HI assays, an individual's HI titre before exposure to influenza virus has a strong inverse relationship with their risk of infection (Hobson et al. 1972; Coudeville et al. 2010), and the HI assay remains the most appropriate correlate of protection available today (Cox 2013).

Neutralisation assays are considered to be functional *assays, compared to HIs which are considered* binding *assays.*

1.3.2 *Antigenic cartography*

Immunological assays generate tables containing serial dilutions which can be difficult to interpret for experts and non-experts alike, especially when the tables contain more than a handful of antigens and antisera. Antigenic cartography aims to represent data from immunological assays in a way that is simple to interpret and that enables a deeper understanding of the underlying data, whilst simultaneously averaging out noise from multiple measurements to generate a more

precise quantitative output (Smith et al. 2004). Antigens and antisera are represented as points in a Euclidean space in the same way that the positions of towns are represented by points on geographical maps. A titre between an antigen and antisera is conceptualised as a measurement of the ideal antigenic distance between the antigen and antisera. A high titre represents an antiserum that is able to block the agglutination of the antigen it is measured against very effectively; the antigen and antisera are deemed antigenically similar, and they are therefore positioned closely in antigenic space. Conversely, lower titres represent antisera that do not block the agglutination of an antigen well, and therefore these antigens and antisera are placed further apart. The problem becomes: what are the positions of antisera and antigens that faithfully represent the titres in a table? This problem can be posed as a computational optimisation of an error function.

ERROR FUNCTION The error function defines how well a particular configuration of antigens and antisera represent target distances derived from a table. A configuration with the lowest error (or *stress*) is then searched for. The error for any configuration is:

$$e = \sum_{ij} (\mathbf{D}_{ij} - \mathbf{d}_{ij})^2 \tag{1.1}$$

Where $\mathbf{D}_{ij}$ are the target distances derived from the table and $\mathbf{d}_{ij}$ are the map distances based on the current configuration of antigens and antisera. Specifically, $\mathbf{d}_{ij}$ is the distance in p dimensional Euclidean space between antigen, i, and antisera, j, in the configuration. Antigen positions are encoded in a matrix, $\mathbf{X}$, with dimensions (n, p), where n is the number of antigens. Similarly, antisera positions are encoded in a matrix, $\mathbf{Y}$, with dimensions (m, p). $\mathbf{d}_{ij}$ is computed as:

$$\begin{aligned} \mathbf{d}_{ij} &= \sqrt{\sum_{k=1}^{p} (x_k - y_k)^2} \\ \mathbf{X}_i &= (x_{i1}, ..., x_{ip}) \\ \mathbf{Y}_j &= (y_{j1}, ..., y_{jp}) \end{aligned} \tag{1.2}$$

$\mathbf{D}_{ij}$ is the target distance specified by a titration, $\mathbf{T}_{ij}$, between antigen i, and antisera j:

$$\begin{aligned} \mathbf{D}_{ij} &= b_j - \log_2(\mathbf{T}_{ij}) \\ b_j &= \log_2(max(\mathbf{T}_j)) \end{aligned} \tag{1.3}$$

Titres represent serial two-fold dilutions, so possible readouts are uniformly spaced on a $\log_2$ scale. The initial dilution of antisera is often $\frac{1}{10}$, so that possible measurements in an assay containing n serial dilutions would be: $(\frac{1}{10}, \frac{1}{20}, \frac{1}{40}, ..., \frac{1}{10 \times 2^n})$. Inputs for antigenic cartography are the reciprocals of dilutions, i.e. $(10, 20, 40, ..., 10 \times 2^n)$.

COLUMN BASIS b_j sets the titration that represents a target distance of zero (b_j is also referred to as the *column basis*). If the maximum titration for an antiserum, $max(\mathbf{T}_j) = 1280$, and the titre of interest, $\mathbf{T}_{ij} = 1280$, then the target distance, $\mathbf{D}_{ij} = 0$. For an antiserum with the same maximum titre, if the titre of interest is one serial dilution lower, $\mathbf{T}_{ij} = 640$, then the target distance $\mathbf{D}_{ij} = 1$. The effect of Equation 1.3 is that one two-fold dilution difference in a titration represents a single unit difference in target distance.

Sometimes a threshold is applied, such that b_j cannot exceed a particular value. This can be useful if, for example, there is one antigen in the data that has very high titres to the majority of antisera (a high avidity antigen). An artificial maximum titre can be used to prevent such a strain distorting the majority of target distances. Likewise, sometimes it is preferable to set a minimum for the column basis, so that an antiserum with uncharacteristically low titres does not attract antigens just by virtue of having a low maximum titre.

Bayesian implementations of antigenic cartography draw values of b_j during sampling of the posterior distribution of model parameters (Bedford et al. 2014). Judging a suitable method for handling column bases can be determined by comparing map errors, and inspecting scatter plots of target distance against map distance (for a perfect map, all points in this scatter would lie on the line $y = x$).

THRESHOLD VALUES Antisera may bind so weakly to test virus such that even the most concentrated dilution does not prevent virus agglutination. Such titres are referred to as being below threshold. In an HI assay, if even a concentration of $\frac{1}{10}$ is unable to prevent virus agglutination, then the result would be recorded as $< \frac{1}{10}$. Below threshold values are handled such that they only contribute to the error if the map distance is lower than the table distance minus one. See (Smith et al. 2004) for details.

OPTIMISATION Optimal configurations of antigens and antisera, $\mathbf{X}$ and $\mathbf{Y}$, are searched for by passing the error function and its first derivative to a gradient descent algorithm such as L-BFGS (Liu and Nocedal 1989). Multiple optimisations using different randomly generated starting configurations are conducted in an attempt to find globally optimal solutions.

APPLICATIONS OF ANTIGENIC CARTOGRAPHY Antigenic cartography was initially applied to seasonal A(H3N2) influenza viruses isolated from humans (Smith et al. 2004). The map enabled deep quantitative insight into the antigenic evolution of these viruses over a 35-year period (Figure 1.4). Most notably, it demonstrated that the antigenic evolution of these viruses is punctuated. The map is characterised by clusters of viruses, that evolved stepwise from one another.

This is in contrast to the gradual genetic evolution of these viruses, which immediately suggests that certain genetic changes must cause a larger antigenic effect than others. Indeed, the authors showed, thanks to a natural experiment whereby a pair of strains differed by a single amino acid substitution, but were members of different antigenic clusters, that a single amino acid change (N145K) was able to cause the transition from the BE92 cluster to the WU95 cluster (Figure 1.4).

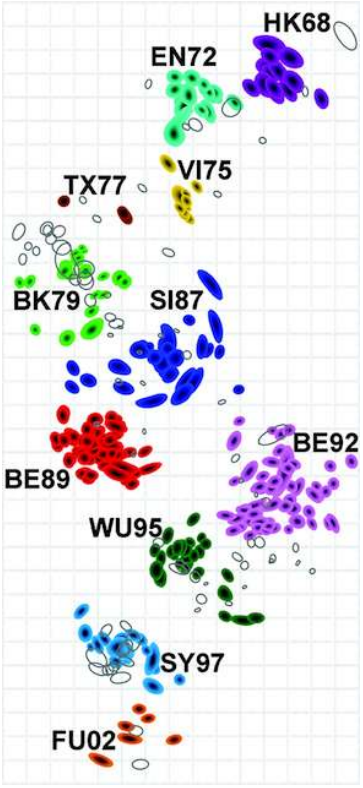


Figure 1.4: Antigenic map of human A(H3N2) influenza viruses isolated between 1968 and 2002 (Smith et al. 2004). Antigens (coloured shapes) and antisera (open shapes) are positioned such that distances between them represent measured titres. The shape of each antigen and antiserum point shows the area that a point can be moved such that map stress does not exceed a threshold. Different colours show 11 clusters of antigens identified by a k-means algorithm. Cluster abbreviations derive from a vaccine strain used in each cluster: HK68, Hong Kong 1968; EN72, England 1972; VI75, Victoria 1975; TX77, Texas 1977; BK79, Bangkok 1979; SI87, Sichuan 1987; BE89, Beijing 1989; BE92, Beijing 1992; WU95, Wuhan 1995; SY97, Sydney 1997; FU02, Fujian 2002.

Assumptions made in antigenic cartography mean its use is not restricted to use with only human A(H3N2) seasonal influenza viruses. Since the initial publication in 2004, antigenic cartography has been widely adopted to address questions in virology and evolutionary biology in different types of influenza viruses including: swine A(H1N1), (Nfon et al. 2011); swine A(H3N2), (Jong et al. 2007; Lewis et al. 2014;

Abente et al. 2016); human A(H1N1), (Garten et al. 2009); human A(H2N2), (Linster et al. 2019); human A(H3N2), (Fonville et al. 2014; Fonville et al. 2016; Kawakami et al. 2019; Russell et al. 2008b); avian A(H5N1), (Koel et al. 2014; Thi Nguyen et al. 2018); and equine A(H3N8), (Lewis et al. 2011). Antigenic cartography is routinely used by the WHO to inform strain choice for influenza vaccines in all major types and subtypes that infect humans (Russell et al. 2008a). Antigenic cartography has also been applied to other antigenically variable pathogens such as Dengue virus (Katzelnick et al. 2015).

1.3.3 *Practicalities of measuring antigenic phenotypes*

Measuring antigenic phenotypes of wild influenza viruses can be hampered by: the fundamental need to passage viruses before they can be assayed, variable agglutination behaviour of some influenza viruses impacting the reliability of assays, and use of non-human animal models to asses antigenic difference.

PASSAGING Influenza viruses in clinical samples cannot be directly assayed because the sample may contain a low concentration of virus, and will contain additional constituents that may interfere with assays. Virus particles can be purified, and their quantity increased, by growing the virus in a cell line, which is known as passaging (Figure 1.5). However, every time a virus is passaged the genome must be copied giving opportunity for mutations to arise either by genetic drift or adaptation to growth in the cells (Widjaja et al. 2006). Of course, genetic mutations may impact phenotypes of interest. The only way to check whether passaging has introduced mutations is to directly sequence the clinical isolate and the passaged virus.

In 1935, it was shown that human influenza viruses could be grown in embryonated chicken eggs (Burnet 1940). Ever since, eggs have been used as a simple and cost effective method for growing influenza viruses. The first influenza vaccine, which was licensed in the U.S.A. in 1945 was grown in chicken eggs (Francis et al. 1945), and many influenza vaccines are still manufactured in chicken eggs today (Krammer 2019). However, molecular adaptation to growth in eggs can change antigenic properties (Chen et al. 2010). To generate antigenic maps of wild type viruses particular care is required to conduct minimal passaging, and avoid passaging in chicken eggs (Smith et al. 2004).

AGGLUTINATION The HI assay relies on the ability of influenza virus particles to agglutinate red blood cells in order to measure the degree to which antibodies can bind virus, and block the agglutination. The agglutination behaviour itself is not directly related to the antigenic phenotype: an antibody may exist which neutralises two

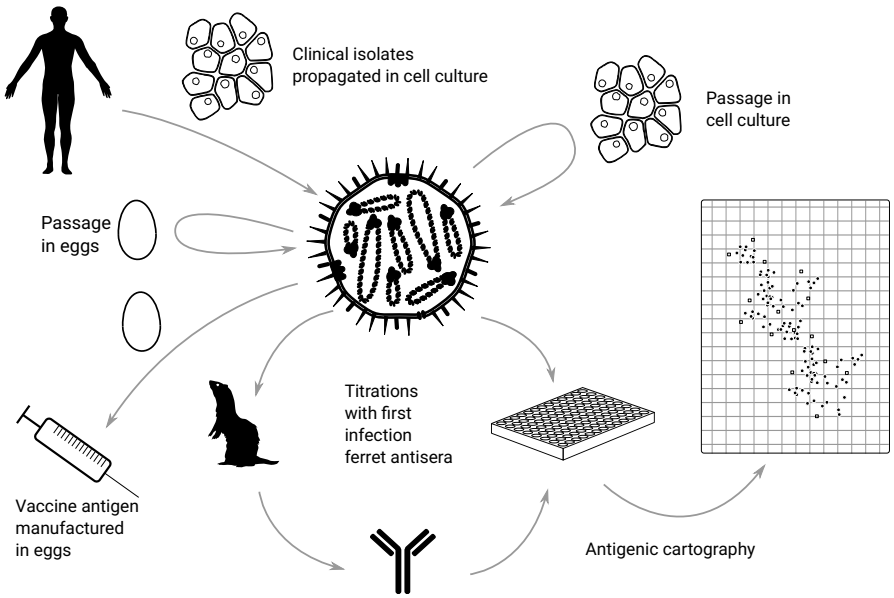


Figure 1.5: Processes in-between virus isolation and antigenic characterisation. This schematic diagram illustrates processes that may occur before antigens and antisera are displayed in an antigenic map. Clinical samples are initially propagated in mammalian cell lines or eggs. Viruses can be subsequently further passaged in mammalian cell lines or eggs. Antisera used in antigenic cartography normally derives from first-infection ferrets.

different viruses equally well, one of which does agglutinate red blood cells, one of which does not. In this case, the HI assay would be unable to be used to measure antigenic properties of the virus which does not agglutinate red blood cells.

Human A(H3N2) influenza viruses isolated after the 1968 pandemic show high variation in binding affinity to different glycans, with seemingly little consequence for virulence or transmission (Lin et al. 2012; Gulati et al. 2013). In the mid 1990s human A(H3N2) influenza viruses stopped binding to chicken red blood cells that were used in HA assays, and a switch to the use of turkey red blood cells was recommended (Gulati et al. 2013). Similarly, in the 2000s the avidity of A(H3N2) influenza viruses to turkey red blood cells dropped, and a switch to guinea pig red blood cells was recommended (Barr et al. 2010; Gulati et al. 2013). Changes in virus avidity to the red blood cells used in HI assays has impeded efforts to conduct comprehensive assays containing representative human A(H3N2) strains from all years from 2002 onwards.

FERRETS Ferrets were the first species that human influenza viruses were shown to be able to replicate in (Smith et al. 1933), and have become a commonly used animal model to study influenza virus infection. Ferrets exhibit many of the same disease symptoms as

humans and influenza viruses can transmit efficiently between ferrets (Oh and Hurt 2016). One factor that affects host tropism in influenza viruses is the ability of the HA to bind glycans with different linkages between the terminal sialic acid and adjacent carbohydrate ring. Avian viruses preferentially bind α-2,3 linked sialic acids, whereas human viruses preferentially bind α-2,6 linked sialic acids (Connor et al. 1994). Ferrets posses α-2,6 linked sialic acids in their upper respiratory tracts, which may increase their susceptibility to human influenza viruses (Jia et al. 2014).

First infection ferret antisera are routinely used to measure the antigenic properties of human influenza viruses (Figure 1.5), making the implicit assumption that the human and ferret immune response to influenza virus infection are comparable. Several pieces of evidence suggest this assumption is broadly justified. Ferret antisera show similar patterns of titres to swine and avian antisera when measured against swine A(H3N2) and avian A(H5N1) strains respectively (Jong et al. 2007; Koel et al. 2014). This suggests that generally the immune response of these three vertebrates is similar, and by extension that ferrets may also be a good model for the human immune response. To be well characterised, antisera is drawn a specific length of time after infection with a known strain. It is unethical to give a human its first influenza infection, making it difficult to test in a well controlled manner whether human and ferret first responses to influenza infection are similar. Nevertheless, highly reactive antisera from children aged between nine and 24 months shows similar patterns to first infection ferret antisera, and generates similar antigenic maps (Fonville et al. 2016). There are however some well documented cases of antigenic differences that are measurable using human antisera but not using ferret antisera (Linderman et al. 2014).

1.4 ANTIGENIC EVOLUTION

As outlined in Subsection 1.1.4, influenza pandemics occur infrequently but when they do occur can infect between 20 % to 40 % of the global population in a year (Taubenberger et al. 2001). The initial wave of infection establishes immunity in the host population which reduces the virulence of subsequent infections, and acts as a selection pressure for antigenic novelty in the virus population. A cyclical feedback loop then ensues whereby virus antigenic evolution drives updated host immunity, and updated host immunity acts as a selection pressure for additional antigenic change (Smith et al. 2004; Nelson and Holmes 2007).

1.4.1 *What governs the antigenic evolution of seasonal influenza viruses?*

Influenza viruses have among the fastest rates of molecular evolution observed in nature (Duffy et al. 2008). The first measurement of the rate of molecular change in influenza A viruses estimated a value of over 7.3×10^{-5} mutations per base, per genome replication (Drake 1993). Given this rate of molecular evolution, one might also expect antigenic evolution to occur rapidly. However, in the last 51 years (1968-2019) there have been only 16 antigenic clusters of human seasonal A(H3N2) influenza viruses. This equates to a rate of approximately 0.31 clusters per year, or 0.29 cluster transitions per year.

For comparison, this value in humans is three orders of magnitude slower: 2.5×10^{-8} (Nachman and Crowell 2000).

In 2014 the amino acid substitutions responsible for cluster transitions that occurred between 1968 and 2002 were elucidated (Koel et al. 2014). One surprising aspect of these results was that most cluster transitions (seven out of ten) were caused by single amino acid substitutions. Furthermore, all but one of the substitutions identified are possible by mutating only a single nucleotide in a codon. Simulations suggest that all possible single nucleotide mutant variants are sampled in the virus progeny in every single influenza virus infection (Russell et al. 2012). The simple molecular architecture of cluster transitions, coupled with the high mutability of influenza viruses begs the question: Why don't new antigenic clusters evolve as soon as even slight immunity to previous antigenic variants develops? More simply: Why is antigenic evolution so slow? Several hypotheses have been proposed to answer this question, which are not necessarily mutually exclusive.

The Koel et al. (2014) and similar work in other subtypes is reviewed in detail in Chapter 4.

NEUTRAL NETWORKS Neutral network theory posits that in fact the molecular basis of antigenic change is not as simple as that presented above. Although single substitutions cause antigenic change, they only do so in specific genetic backgrounds, either because a compensatory substitution is required to offset some detrimental functional or structural impact, or because the antigenic impact of a substitution is only observed in a specific genetic background (Koelle et al. 2006). The time delay is therefore explained as time spent by the virus population performing a random walk on a network of evolutionary neutral genotypes until one is sampled that permits the cluster transition substitution.

FITNESS EXCHANGE The fitness exchange hypothesis suggests that cluster transition substitutions negatively impact virus replication in some way, and that the substitution is only selected when its negative effects are outweighed by the positive effect of its antigenic escape. It is useful to partition the fitness of virus strains into two components: *Extrinsic fitness* relates to the degree to which a virus evades prior immunity in the host population. A virus in a region of antigenic space where many strains have already circulated would have low

extrinsic fitness because individuals in the host population would be able to mount effective immune responses. A virus in a novel region of antigenic space experiences higher extrinsic fitness because it has a larger pool of susceptible hosts. *Intrinsic fitness* encompasses all other aspects of influenza virus replication.

Of key importance to the fitness exchange mechanism is how intrinsic and extrinsic fitness change over the course of an antigenic cluster. When a new antigenic cluster of strains begins to circulate, there is little extrinsic fitness difference between viruses in the new cluster, and viruses in an even more advanced cluster. However, over time, immunity to the current cluster develops in the host population, meaning that the extrinsic fitness of strains in the advanced cluster grows. The fitness exchange mechanism predicts that cluster transition substitutions occur only when the reduction in intrinsic fitness of cluster transition substitutions are offset by the increase in extrinsic fitness.

There is theoretical reason to suspect an interrelationship between extrinsic and intrinsic fitness of substitutions. Cluster transition substitutions in human seasonal A(H_3N_2) influenza viruses all occur on the periphery of the HA RBS (Koel et al. 2013). Therefore, substitutions that cause antigenic escape may also impact HA binding, and negatively interfere with replication.

STOCHASTICITY The timing of cluster transition substitutions may contain large stochastic components at both within and between host level.

Although in simulations all possible single nucleotide variants are sampled in every single influenza virus infection (Russell et al. 2012), the same work also suggests that final progeny is dominated by relatively few variants that arise early in the infection cycle. So, even though many variants are generated, the vast majority do not get transmitted, and therefore do not have a chance of being selected in nature.

Then, even if an antigenically novel variant were transmitted to a new host, stochastic effects at the host population level may prevent that variant from spreading to fixation. For example, the variant may evolve in a region with poor travel connections to other regions and exhaust its local pool of susceptible hosts before spreading further.

1.5 VACCINES

Krammer and Palese (2015) is an excellent review of influenza virus vaccine technology.

The first influenza virus vaccines were developed in the 1940s and termed whole-virus inactivated vaccines. Virions were propagated in eggs and inactivated chemically (Burnet 1940). Some manufactures started using Madin-Darby Canine Kidney (MDCK) cells and other mammalian cell lines to reduce unwanted by-products from egg-

based manufacture such as ovalbumin (the main protein in egg white). A second reason was to prevent the selection of genetic adaptations to growth in eggs, which can alter antigenic properties (Chen et al. 2010). Not all egg adaptation mutations cause antigenic change; some increase growth rate which is useful for vaccine manufacture (Barman et al. 2015).

Nowadays there are several types of Inactivated Influenza Vaccines (IIVs). It is common to treat grown vaccine virus with detergent which disrupts viral membranes and frees viral proteins to generate split vaccines. Additional purification can also occur to extract or remove specific protein domains to generate subunit vaccines. The development of split and subunit vaccines aimed to increase the relative proportion of the HA globular head in the vaccine, as these vaccines function by eliciting the production of HA specific antibodies (Wood 1998), whilst also decreasing other reactogenic components of the virus. Current IIVs contain strains from three or four different subtypes. Trivalent vaccines contain one A(H3N2) strain, one A(H1N1) strain and one B virus strain. Quadrivalent vaccines contain strains from both influenza B virus lineages and the A(H3N2) and A(H1N1) strains. Vaccines are administered either intradermally, or intramuscularly and contain 15 µg of HA from each subtype component. Quantities of other viral proteins are not standardised.

Growth of vaccine antigen in either eggs of mammalian cell lines has several undesirable side effects: 1) high growth reassortant virus which may be genetically and antigenically different from the selected vaccine strain must be identified and used; 2) high level bio-containment is required because the process involves growth of actual influenza viruses; 3) toxic chemicals are required to inactivate viruses; and 4) concentrations of by-products such as ovalbumin, and antibiotics used to keep the growth medium sterile are difficult to control, or require additional downstream purification. The use of expression systems to generate recombinant protein vaccine antigen circumvents all of these issues. HA vaccine antigen used in Flublok is expressed in Sf9 cells from the fall armyworm moth (*Spodoptera frugiperda*), and is licensed for use in the U.S.A. (Cox et al. 2015).

Live Attenuated Influenza Vaccines (LAIVs) utilise cold adapted strains that can replicate effectively in the upper respiratory tract but not in the lungs. LAIVs are highly effective in young children (Belshe et al. 2007).

Many novel vaccine technologies are currently in development or in clinical trials: ΔNS1 second generation attenuated vaccines which lack the NS1 gene; RNA-based and DNA-based vaccines whereby coding sequences are administered and host cellular machinery generates and presents antigen to immune cells; virus like particles; and intranasal whole inactivated vaccines (Krammer 2019).

1.5.1 *Vaccine strain selection*

Influenza viruses can spread quickly worldwide (Russell et al. 2008b; Pybus et al. 2015). The importance of a global context for influenza virus evolution was quickly acknowledged: in 1952 the WHO established the Global Influenza Surveillance and Response System (GISRS) to facilitate global collaboration in sharing influenza viruses, data and research (Payne 1953). Currently, the GISRS comprises 115 National Influenza Centres (NICs) in WHO member states which share surveillance samples and data. Five Collaborating Centres in Melbourne, Australia; Beijing, China; Tokyo, Japan; London, U.K.; and Atlanta, U.S.A., conduct centralised immunological assays and genetic sequencing on the surveillance samples from NICs. These data are collated by the WHO Collaborating Centre in Cambridge, U.K., and processed to generate antigenic maps and phylogenetic trees. A WHO committee meets twice a year to evaluate these analyses and recommend a suitable vaccine strain for each human subtype. After strain selection, roughly eight months is required for vaccine manufacture, testing and distribution. This *deployment lag* means that strain selection must occur eight months before the onset of the influenza season in each hemisphere. In the Northern hemisphere the influenza season normally falls between October and May, meaning that the vaccine strain selection must occur in February. The Southern hemisphere season falls between April and November, and vaccine strain selection occurs in September.

A sixth Collaborating Centre in Memphis, U.S.A., specialises in animal influenza viruses.

Current vaccines are licensed such that they must contain a strain that is antigenically representative of the currently circulating viruses. Influenza virus antigenic evolution is punctuated, so at a high level this selection process can be summarised as selecting a strain that is in the centre of the currently circulating cluster of viruses (Figure 1.4). This system selects well-matched strains (Russell et al. 2008a). However, if an antigenic cluster transition occurs after vaccine strain selection, but before the onset of the influenza season then the subsequent season can still be antigenically mismatched. Theoretically, reducing deployment lag would reduce the incidence of mismatched vaccines (Figure 1.6), however, vaccine deployment is highly optimised and there is little room for improvement.

Surveillance lag is the delay in time between the evolution of a new antigenic cluster in nature and its detection in surveillance such that it can inform vaccine strain selection. Surveillance lag should be reduced as much as possible, in order to have the most up to date data for vaccine strain selection. However, even if there were zero surveillance lag, cluster transitions could still occur during the more substantial period of deployment lag (Figure 1.7b). Therefore, the current vaccine strain selection procedure will not be able to avoid antigenic mismatch between circulating strains and vaccines due to

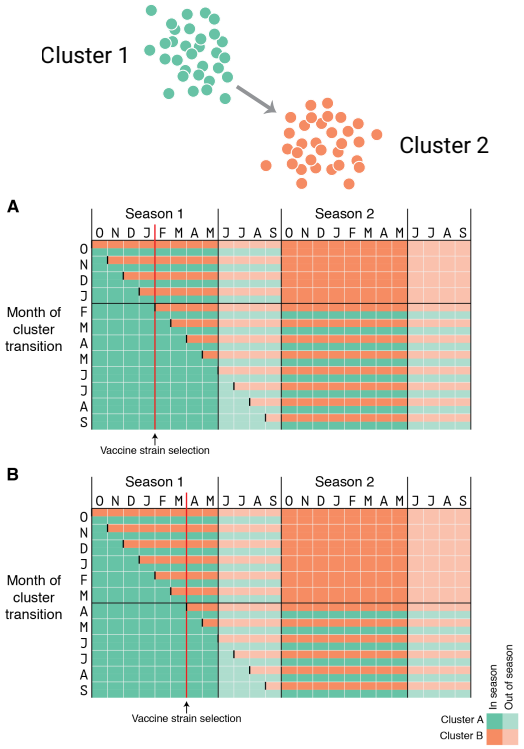
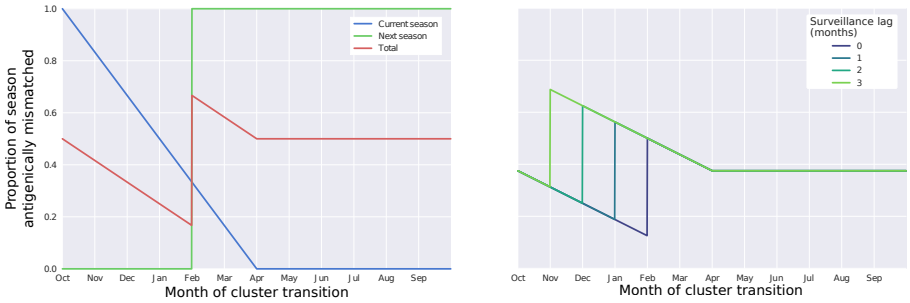


Figure 1.6: The effect of cluster transition timing and vaccine strain selection timing on antigenic mismatch. Consider viruses evolving from antigenic cluster one (green) to cluster two (orange). The top, and bottom, of each cell indicates the cluster of circulating viruses, and the vaccine strain, respectively. When the vaccine is matched, the cell appears as a single colour; when it is mismatched, the cell contains two colours. Each row demonstrates cluster transitions occurring at the start of a different month, starting in October of Season one. If the cluster transition occurs before vaccine strain selection, then the vaccine strain can be updated to match in the subsequent season. Table A illustrates the scenario where vaccine strain selection occurs at the start of February, i.e. matching the case in the Northern hemisphere currently. In Table B vaccine strain selection occurs two months later, at the start of April. This results in a higher number of matched cells in season two.

antigenic cluster transitions occurring during vaccine manufacture and deployment.

1.5.2 *Antigenically advanced influenza vaccines*

One way to overcome the antigenic mismatch that the current vaccine strain selection procedure cannot avoid is to use antigenically advanced influenza vaccines. Here, the aim is to vaccinate individuals with strains that do not yet exist naturally, but that will exist in the future. In doing so, individuals acquire immunity to antigenic variants



(a) Effect of cluster transition timing on mismatch in current and next season

(b) Varying surveillance lag

Figure 1.7: Impact of cluster transition timing and surveillance lag on antigenic mismatch. The month that the cluster transition occurs is plotted against the proportion of a season that is antigenically mismatched as a result. (a) In the current season (blue), the later the cluster transition occurs, the less the current season is mismatched. The next season (green) is either entirely matched or mismatched, whether the cluster transition occurred before or after February when vaccine strain selection occurs. The combined proportion of the current and next season mismatched is shown in red. (b) Each trace shows the proportion of the combined (current and next) season mismatched for different lengths of surveillance lag.

that will circulate in the future. Clearly, this strategy relies on the ability to know ahead of time what antigenic variants will circulate.

1.6 OUTLINE OF THESIS

The scope of this thesis is to (1) quantitatively determine the impact of antigenic mismatch on vaccine effectiveness is and (2) test how predictable the antigenic evolution of seasonal A(H3N2) influenza viruses are.

In Chapter 3 I investigate what the quantitative relationship is between antigenic mismatch and Vaccine Effectiveness (VE), using empirical measurements of mismatch derived from antigenic maps and VE estimates derived from the literature.

In Chapter 4 I review what is currently known about the genetic basis for major antigenic change in influenza viruses. Furthermore, I present a new quantitative method for testing which amino acids are associated with major antigenic change based on a LMM framework adapted from quantitative genetics. The LMM methods are detailed in a second introductory chapter: Chapter 2. I also present analyses consisting of identifying virus strains that comprise natural experiments that test the impact of amino acid substitutions on antigenic change. I show that these molecular changes contain a high degree of non-randomness, and suggest that these patterns are associated with

the biological mechanisms that underlie the evolution of antigenic escape substitutions.

In Chapter 5 I develop a ranking framework that uses the patterns identified in Chapter 4 to test how predictable future antigenic cluster transitions would be using these patterns alone. In Chapter 6 I discuss how these rankings could be used with additional experimentally derived virus fitness data to substantially narrow down candidate viruses that may circulate in the future. I go on to refine the concept of antigenically advanced influenza vaccines and pre-emptive vaccine updates through the lens of *immunity management*, and argue that this strategy could be used in adults to ameliorate low VE caused by antigenic mismatch.

2

INTRODUCTION TO LINEAR MIXED MODELS FOR ANTIGENIC PHENOTYPES

CHAPTER OUTLINE Linear Mixed Models (LMMs) are used in quantitative genetics to identify genetic markers associated with phenotypes. LMMs are a work horse in this field because they provide a framework to conduct statistical tests in large datasets with complex patterns of relatedness between samples. These properties make LMMs attractive for applying to questions in influenza virus biology. Here, I briefly introduce LMMs and their application in quantitative genetics, and explain how I have developed LMMs to antigenic coordinates and amino acid sequences, how association tests are conducted, and genotype to phenotype mapping using the Best Linear Unbiased Predictor (BLUP).

This chapter introduces methods used elsewhere in this thesis. In Chapter 3 I use BLUP to predict antigenic coordinates of strains that have been sequenced, but not titrated, to increase the size of datasets from which mismatch distributions are derived. In Chapter 4 I use a LMM to statistically test whether molecular changes are associated with antigenic cluster transitions. Finally, BLUP is the basis for a novel post-processing step, map sanitising, that I present in Appendix A.

2.1 WHAT IS A MIXED MODEL?

Before tackling Multivariate Linear Mixed Models (mvLMMs), and their application in quantitative genetics, I will outline what the features of a simple mixed model are, and why they are used. Simply put, a mixed model contains both fixed effects, and random effects. A fixed effect is a component in a model for which a parameter is estimated that corresponds to a specific experimental treatment. This could be as simple as estimating means for different groups in data, or estimating the coefficient of a regression slope. In contrast, a random effect models a variable by estimating the parameters of a probability distribution.

I will illustrate the difference between random and fixed effects using a simple dataset consisting of measurements of the time taken for ultrasonic waves to travel the length of six randomly selected railway rails (Pinheiro and Bates 2000; Devore 2000). Concepts taken from this basic example will apply to more complex, and higher dimensional models introduced later on.

Ultrasonic travel time is used to asses degradation in rail microstructure.

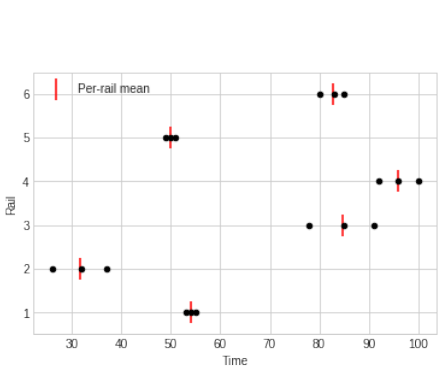
In this dataset, each rail was measured three times, so the data are naturally grouped according to the rail the measurement was taken from. One approach is to model individual rails as fixed effects:

$$\begin{aligned} y_{ij} &= \beta_i + \epsilon_{ij} \\ \epsilon_{ij} &\sim \mathcal{N}(0, \sigma^2) \end{aligned} \tag{2.1}$$

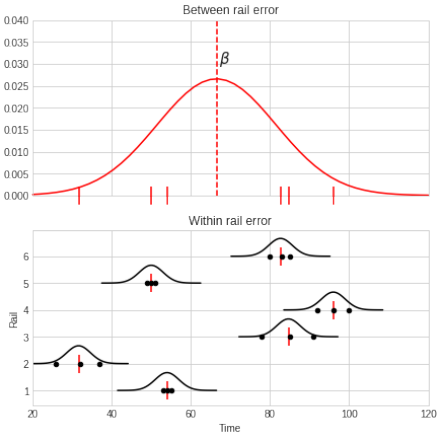
where y_{ij} is the travel time for the j^{th} repeat of the i^{th} rail, ϵ_{ij} captures the per measurement error and β_i is the mean travel time within rail i:

$$\beta_i = \frac{1}{n} \sum_{j}^{n} y_{ij} \tag{2.2}$$

This models the specific sample of rails sampled by computing the mean travel time in each rail. Measurement error is assumed to be drawn from one normal distribution for all rails, with variance σ^2. It estimates the central tendency of the travel time for each rail in the sample, but does not provide insight on the between-rail variability or the wider population that the rails were sampled from. The number of parameters increases linearly with the number of rails sampled.



(a) Fixed effect model



(b) Random effect model

Figure 2.1: Rails dataset modelled using fixed and random effects. In the fixed effect model, one parameter is estimated for each rail: the mean travel time. In the random effect model, variance is decomposed into between rail (top) and within rail (bottom) components.

Alternatively, each rail could be modelled as a random effect (Figure 2.1b). The focus changes from estimating parameters that describe the particular rails sampled, to describing the data in terms of the variance associated within a rail, and between rails:

$$\begin{aligned} y_{ij} &= \beta + b_i + \epsilon_{ij} \\ b_i &\sim \mathcal{N}(0, \sigma_b^2) \\ \epsilon_{ij} &\sim \mathcal{N}(0, \sigma^2) \end{aligned} \tag{2.3}$$

β is now the grand mean of the travel time in all rails, and b_i is a per-rail random effect. b_i is modelled using a normal distribution, with mean zero, and variance σ_b^2, which captures between-rail variability. b_i are *random*, because they are associated with experimental units (rails) drawn at random from a wider population of rails. b_i are also *effects* because they alter the measurements for rail i by a particular amount away from the population mean, β.

In contrast to the fixed effect treatment, per-rail effect sizes are not directly estimated. Rather, the total variance in the data is decomposed into one component that captures within rail variability (σ^2) and another component that captures between rail variability (σ_b^2). An important consequence of this is that as the number of samples in the data grows, the number of parameters being estimated stays constant, and degrees of freedom do not grow. This is not the case in the fixed effect model, where each rail sampled requires an additional β_i be estimated.

2.2 MIXED MODELS IN QUANTITATIVE GENETICS

A foremost aim in genetics is to detect markers responsible for phenotypes. It was noted over 25 years ago that genetic samples are often non-independent due to varying degrees of relatedness between individuals which invalidates assumptions of most statistical tests (Spielman et al. 1993). Structure in biological populations is common. In humans, the average South American is likely to share more genetic markers with another South American than with a European. This poses a problem for association studies; if a phenotype is more prevalent in a group because of a shared causative locus, the group probably also shares many additional markers. Therefore, it can be difficult to distinguish causative markers from those that are just frequent in the group (Balding 2006). Whilst other methods exist to adjust for population stratification in Genome Wide Association Studies (GWAS) (Devlin and Roeder 1999; Pritchard et al. 2000; Price et al. 2006), mixed model approaches have come to dominate the field due to their ability to correct for relatedness at different scales (Kang et al. 2010; Zhang et al. 2010; Korte et al. 2012; Zhou and Stephens 2012; Lippert et al. 2011).

LMMs decompose a measured phenotype, $\mathbf{y}$, into the sum of several components:

$$\mathbf{y} = \beta\mathbf{x} + \boldsymbol{\alpha}\mathbf{F} + \mathbf{g} + \boldsymbol{\psi} \tag{2.4}$$

Mixed models in quantitative genetics are normally linear *because the fixed effects are included as linear terms.*

A genetic marker, usually a single Single Nucleotide Polymorphism (SNP), is represented as a fixed effect, $\beta\mathbf{x}$. β is the effect size, and $\mathbf{x}$ encodes the SNP under a particular inheritance model (also referred to as the genetic profile). In diploid organisms, a dosage inheritance model would encode a homozygote lacking the SNP as 0, a heterozygote as 1, and a homozygote carrier as 2. Other covariates, $\mathbf{F}$, that may influence

the phenotype can also be included as fixed effects, with effect size α. Typically, covariates include factors such as age, sex and weight. Genetic kinship is included as a random effect, $\mathbf{g}$, which expresses the degree to which phenotypes of two individuals should covary based on their relatedness. This is achieved by treating $\mathbf{g}$ as a multivariate normally distributed random variable:

$$\mathbf{g} \sim \mathcal{N}(0, \sigma_g^2 \mathbf{R}) \tag{2.5}$$

with mean, 0, and covariance matrix, $\sigma_g^2 \mathbf{R}$. $\mathbf{R}$ measures the genetic relatedness. It is the covariance matrix of the entire SNP matrix, $\mathbf{X}$:

$$\mathbf{R} = \frac{1}{S} \mathbf{X}\mathbf{X}^T \tag{2.6}$$

With N individuals, and S SNPs, $\mathbf{X}$ has dimensions (N, S). Therefore, $\mathbf{R}$ is an (N, N) symmetric matrix where cell ij stores the genetic similarity between individuals i and j. Residual variation not captured by the other components is modelled as:

$$\boldsymbol{\psi} \sim \mathcal{N}(0, \sigma_n^2 \mathbf{I}_N) \tag{2.7}$$

where $\mathbf{I}_N$ is the identity matrix. The scalars σ_g^2 and σ_n^2 modulate the contribution of the kinship, and error components respectively.

2.2.1 *The essence of LMMs*

The key feature of a LMM is modelling genetic kinship using a random effect. This random effect encompasses the degree to which phenotypes covary, based on overall genetic similarity. In the rail example, groups are discreet, meaning that the measurement of a single sample belongs to a single group. The degree to which a rail deviates from the population mean is decomposed into a between-group term, and within-group term, and the between rail variability is measured by a single scalar, σ_b^2 (Equation 2.3). $\mathbf{R}$ aims to do something similar, but on a finer scale. Individuals that covary perfectly with each other will have the maximal covariance value S/N, whereas individuals who share no genetic markers whatsoever will have a value of 0. $\mathbf{R}$ therefore flexibly models all degrees of relatedness on an individual by individual basis.

2.2.2 *Multivariate LMMs*

The first LMMs used in GWAS were applied to datasets containing single traits (Kang et al. 2010; Zhang et al. 2010; Zhou and Stephens 2012; Lippert et al. 2011). The motivation behind the genetic random effect in LMMs is to appropriately handle correlation in the genetic markers of the samples. However, many biological datasets contain multiple

correlated phenotype and some biological phenotypes are inherently multi-dimensional. Conducting multiple separate tests on variables that are correlated will result in a loss of statistical power so it can be more desirable to model such variables jointly. This motivated the development of LMMs that correctly handle multiple traits in the response variable: mvLMMs (Korte et al. 2012; Lippert et al. 2014).

The antigenic diversity of influenza and dengue viruses are good examples of multidimensional biological phenotypes (Smith et al. 2004; Lewis et al. 2011; Katzelnick et al. 2015; Lewis et al. 2014).

The univariate LMM (Equations 2.4-2.7) can be generalised to multiple traits. The mvLMM for P phenotypes, is modelled with a matrix-variate distribution:

$$\begin{aligned} \mathbf{Y} &= \mathbf{x}\boldsymbol{\beta}^{T} + \mathbf{G} + \boldsymbol{\psi} \\ \mathbf{G} &\sim \mathcal{MN}_{N,P}(0, \mathbf{R}, \mathbf{C_g}) \\ \boldsymbol{\psi} &\sim \mathcal{MN}_{N,P}(0, \mathbf{I}_N, \mathbf{C_n}) \end{aligned} \tag{2.8}$$

This differs to the univariate case in three ways: Firstly, the phenotype $\mathbf{Y}$ is an (N, P) matrix and $\boldsymbol{\beta}$ is a length P effect size vector. Secondly, the genetic random effect, $\mathbf{G}$, is modelled using a matrix-variate distribution. $\mathbf{R}$ specifies the individual-to-individual covariance, and $\mathbf{C_g}$ specifies the genetic trait-to-trait covariance. Finally, residual error $\boldsymbol{\psi}$ is also modelled with a matrix-variate distribution with no individual-to-individual covariance, $\mathbf{I}_N$, and $\mathbf{C_n}$ parametrises the trait-to-trait covariance. Covariates can be included in mvLMMs but are omitted from Equation 2.8 because they were not used in this thesis. For further detail on LMMs, mvLMMs, their derivation and parameter estimation, see (Lippert et al. 2014; Casale 2016; Meyer 2017) and the references therein. All analyses in this thesis were conducted using the LIMIX framework for LMM (Lippert et al. 2014).

2.3 ADAPTING LMMS FOR INFLUENZA VIRUS ANTIGENIC PHENOTYPES

mvLMMs lend themselves well to modelling the molecular basis of antigenic phenotypes in influenza viruses. Multiple dimensions are typically required to embed antigenic coordinates and influenza virus populations tend to have high degrees of structure. The degree to which two influenza strains covary genetically, and the antigenic distance between them are negatively correlated (Figure 2.2). Therefore, it seems attractive to account for the degree to which virus strains covary phenotypically based on the degree to which they covary genetically.

2.3.1 *AAPs vs. SNPs*

The main difference between the standard use of LMMs and their use in this thesis is that the independent variables derive from amino acid alignments of the HA gene, rather than microarray-derived data that encode the presence or absence of SNPs among individuals. Dummy

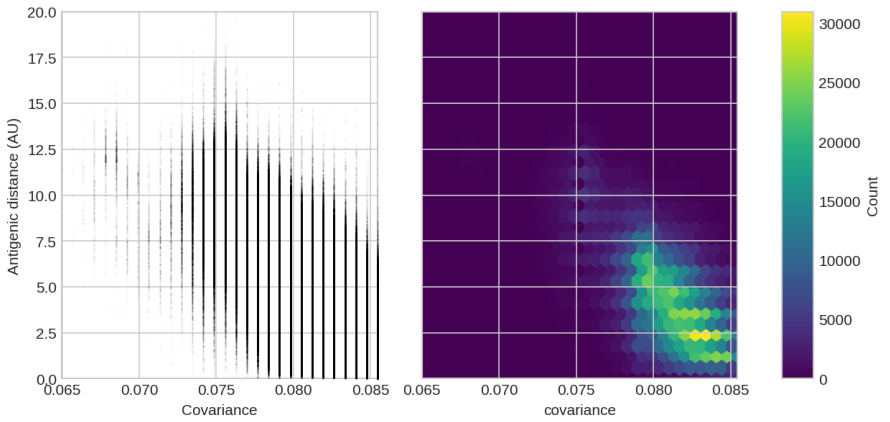


Figure 2.2: Antigenic distance and genetic covariance between pairs of influenza virus strains. There are 2.01×10^6 points in the scatter plot (left), therefore a 2D histogram (right) is useful to show the dense region of the data. Histogram bins are hexagons which tessellate the space. The number of points within each hexagon is shown with a colour scale. This is the Russell et al. (2008b) dataset (see Section 2.4 for details).

variables encode the presence or absence of a particular amino acid at a particular position in an HA alignment (Table 2.1). Invariant dummies provide no information on differences between strains, so are removed. It is not uncommon for multiple dummies to encode identical patterns of presence or absence (e.g. 2K and 3S, Table 2.1B). In these instances there is no way to separate the effects of one Amino Acid Polymorphism (AAP) from the other, so the AAPs are merged. AAPs can also be merged if they encode precisely the opposite pattern of presence and absence from each other. Conceptually, there is no basis to prefer encoding the presence of an AAP as a 0 rather than a 1. Mathematically, using the alternate encoding requires transforming the effect size by a factor of minus one. Merging identical or precisely opposite AAPs decreases the computational burden by decreasing the number of positions that have to be tested in association studies. More importantly, decreasing the number of AAPs decreases the number of statistical tests that have to be conducted. Commonly, p-values are adjusted according to the number of tests that are conducted to reduce false positives. Therefore, decreasing the number of statistical tests reduces the factor by which p-values are adjusted, effectively increasing statistical power.

2.3.2 *Number of tests*

Additional considerations make fitting mvLMMs to influenza datasets practical. mvLMMs have been developed in the GWAS field, where datasets containing 10^4 to 10^5 individuals and 10^7 SNPs are common. Therefore, mvLMMs implementations are computationally effi-

(A) Molecular alignment

Strain	1	2	3	4
A	D	K	S	H
B	D	T	F	H
C	E	T	G	H
D	E	K	S	H

(B) Dummy variables

1D	1E	2K	2T	3S	3F	3G	4H
1	0	1	0	1	0	0	1
1	0	0	1	0	1	0	1
0	1	0	1	0	0	1	1
0	1	1	0	1	0	0	1

(C) Duplicates merged and invariant removed

Strain	D1E	K2T | 3S	3F	3G
A	1	1	0	0
B	1	0	1	0
C	0	0	0	1
D	0	1	0	0

Table 2.1: Representation of protein sequences in LMM. Positions in a molecular alignment *(A)* are converted to dummy variables *(B)* which encode the presence or absence of an amino acid at a position in the alignment. *(C)* Duplicate columns are merged. Columns are considered duplicate if their contents are identical (e.g. 2K and 3S), or if their contents would be identical if inverse coding were used (e.g. 1D and 1E). Invariant columns are also removed (e.g. 4H).

cient and run fast on datasets used in this thesis that typically contain a similar number of individuals, but many fewer genetic markers. Furthermore, the task of inferring molecular changes associated with antigenic change in influenza viruses is substantially more specific than the questions normally addressed using mvLMMs. Not only is the influenza virus genome five orders of magnitude smaller than the human genome, which would itself drastically decrease the number of tests required, but we know that major antigenic change is mediated by the HA gene (Wiley et al. 1981; Wilson and Cox 1990). Further, Koel et al. (2013) used chimeric viruses to show that the antigenicity of A(H3N2) influenza viruses is determined by HA positions 109-301 located in the globular head. It has also been shown in A(H5N1) influenza viruses, that the HA globular head determines antigenicity (Richard, 2017, *pers. commun.*). These data narrow down the HA positions that have to be screened. The higher number of markers that are tested in a typical GWAS requires more stringent methods for correcting for multiple testing, which reduces effect sizes that can be reliably detected.

Conversely, using AAPs rather than SNPs increases the number of tests relative to standard GWAS. AAPs encode amino acid diversity (of which there are 20 in the standard set) whereas SNPs encode nucleotide diversity (of which there are four in standard DNA). $n-1$ dummy variables are required to encode data containing n categories, so, a

single HA position may require as many as 19 AAPs to encode the diversity in a dataset. It would be extremely rare to see such high diversity at a single position. Even if there were five distinct AAPs that had to be tested at all 301 positions in HA1, this would only require running $(301 - 108) \times 5 = 965$ tests. Compared to questions normally addressed using mvLMMs, the applications in this thesis can be more specific, due to the small size of the influenza virus genome, and the large amount of previous work that has pinpointed one region of one gene as the major determinant of the phenotype of interest.

2.3.3 *Association testing*

Association testing aims to formally test the hypothesis that a marker is associated with a phenotype, whilst accounting for confounding factors like population structure, and covariates like age, sex or weight. The association test is framed in typical frequentist form, whereby a null hypothesis ($H_0 : \beta = 0$) is competed against an alternate hypothesis ($H_1 : \beta \neq 0$). False positives occur when the null hypothesis is incorrectly rejected. A threshold, α, is decided *a priori* for an acceptable false positive rate. A p-value is then calculated which measures the probability of observing the given, or more extreme data, under the null hypothesis. The null hypothesis is rejected if p-value $< \alpha$. The log-likelihood ratio test statistic, D, can be used to compare the likelihood of the alternate model, H_1, to the null model, H_0, from which the p-value is derived:

In mvLMM β is a length P vector containing the effect size in each phenotype dimension (Equation 2.8).

$$D = \mathcal{L}(Y|\hat{\beta}, \hat{C_g}, \hat{C_n}) - \mathcal{L}(Y|0, \bar{C_g}, \bar{C_n}) \tag{2.9}$$

$\mathcal{L}(Y|\hat{\beta}, \hat{C_g}, \hat{C_n})$ is the likelihood given the maximum likelihood estimates of the parameters under the alternate model, and $\mathcal{L}(Y|0, \bar{C_g}, \bar{C_n})$ is the likelihood given the maximum likelihood estimates of the parameters under the null model. 2D follows a χ^2 distribution from which a p-value can be derived.

mvLMMs test phenotypes jointly, which can increase statistical power in two ways. Firstly, differences in phenotypes between groups of individuals can be larger in joint dimensions, than individual dimensions (Figure 2.3). In the Figure 2.3 example, the difference between the red and blue individuals is smaller in Y_1 and Y_2 than in the joint combination of Y_1 and Y_2. The joint dimension could be mapped to an individual dimension by a 45° rotation about the origin in either direction. However, this would require knowing *a priori* what particular rotation should be conducted. Moreover, in more complex, and realistic, cases it may not be possible to detect which dimensions are those most strongly associated with particular genetic effects. Secondly, testing all phenotype dimensions in a single analysis reduces the number of tests, and therefore less stringent multiple testing correction has to be conducted.

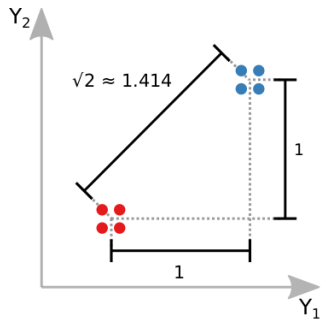


Figure 2.3: Joint analyses can detect larger effect sizes. Phenotypes Y_1 and Y_2 are plotted for eight samples which differ at a single marker, shown with colour. The effect on each individual phenotype of the genetic marker is 1, whereas the effect on the joint phenotypes is $\sqrt{2}$.

2.3.4 *Correcting for multiple tests*

Detecting which out of a set of genetic markers may be associated with a phenotype of interest inherently requires conducting multiple tests. This presents a problem in that running more tests increases the risk of one or more p-values being smaller than α purely by chance. For example if $\alpha = 0.05$ and twenty trials are conducted on a system that generates random p-values, by chance you would expect one trial to have a p-value lower than α. A common way to account for multiple testing is to multiply p-values by a factor equal to the number of tests conducted, known as Bonferroni correction (Bonferroni 1936). In the example, this would mean multiplying each p-value by 20.

Or divide α by the same factor.

For most genetic analyses this would penalise p-values too harshly, due to potentially high degrees of correlation between genetic markers. If in a study of 1000 individuals there were only two markers, and their profiles only differed in one individual, you would expect their p-values and effect sizes to be similar. Intuitively, in this case it seems wrong to scale the p-values by a full factor of two, given that the independent variables in both cases are highly correlated. Instead, p-values are scaled by the effective number of tests, t, which is computed based on correlation among genetic markers. In this thesis t is calculated using the Galwey (2009) method. Adjusted p-values are computed by scaling a raw p-value by a factor of t.

2.3.5 *Per-test phenotype variance decomposition*

A computationally intensive step in GWAS is to decompose the phenotype variance into a genetic and error component. Mathematically, this step is equivalent to estimating σ_g^2 and σ_n^2 in Equations 2.5 and 2.7, and requires having estimated $\mathbf{R}$. In human GWAS effect sizes are small meaning that computing σ_g^2 and σ_n^2 once, and reusing the

values for each test conducted, is a good approximation and saves computation. Single substitutions can have large effects on influenza virus antigenicity (Lewis et al. 2011; Koel et al. 2013; Abente et al. 2016; Koel et al. 2014), meaning this is a bad approximation which may cause p-value deflation. Furthermore, in GWAS studies it is not feasible to recompute $\mathbf{R}$ on a per marker basis. Due to the comparatively small number of markers used in this thesis the phenotype variance can be decomposed for each association test. For each test, the AAP being tested is removed from $\mathbf{X}$, $\mathbf{R}$ is recomputed and σ^2_g and σ^2_n are re-estimated.

2.3.6 *Epistasis*

Epistasis has subtly different meanings in different fields (Lehner 2011). Geneticists often use epistasis to describe the effects of one gene masking another (Bateson 1909). In population genetics, it refers more generally to any deviation from an additive combination of two markers and is also sometimes referred to as an *interaction*. Here, I use the second more general definition.

Testing for interactions is not routinely conducted in GWAS due to the large number of tests required, although the question has received attention (Marchini et al. 2005; Hu et al. 2014). In a study with S markers, $(S-1)^2/2$ additional tests are required to test all pairwise interactions. As previously mentioned, the influenza virus genome is five orders of magnitude smaller than the human genome, and the major determinant of the molecular basis of antigenic change has been honed down to a small region in HA. The relatively small number of markers to screen makes hunting for epistatic effects in the molecular basis of antigenic change feasible. Testing for epistasis can be achieved by extending the mvLMM framework presented above.

Consider two markers, $\mathbf{x_A}$ and $\mathbf{x_B}$. Their interaction, $\mathbf{x_{AB}}$, is the pairwise logical 'and' of their profiles (Table 2.2). We would like to test whether a fixed effect coefficient for $\mathbf{x_{AB}}$ is not equal to zero. Equation 2.8 can be modified so that $\mathbf{x_A}$, $\mathbf{x_B}$ and $\mathbf{x_{AB}}$ are included as fixed effects:

$$\begin{aligned} \mathbf{Y} &= \mathbf{x_A}\boldsymbol{\beta}_\mathbf{A}^\mathbf{T} + \mathbf{x_B}\boldsymbol{\beta}_\mathbf{B}^\mathbf{T} + \mathbf{x_{AB}}\boldsymbol{\beta}_\mathbf{AB}^\mathbf{T} + \mathbf{G} + \boldsymbol{\psi} \\ \mathbf{G} &\sim \mathcal{MN}_{N,P}(0, \mathbf{R}, \mathbf{C_g}) \\ \boldsymbol{\psi} &\sim \mathcal{MN}_{N,P}(0, \mathbf{I}_N, \mathbf{C_n}) \end{aligned} \tag{2.10}$$

The genetic and error random effects components are the same as before, except that when variance is decomposed for each association test, the two markers comprising the interaction term being tested are both omitted when recomputing the Realised Relationship Matrix (RRM). The interaction term is tested by competing a null hypothesis ($H_0 : \boldsymbol{\beta_{AB}} = 0$) against an alternate hypothesis ($H_1 : \boldsymbol{\beta_{AB}} \neq 0$) as described above.

x_A	x_B	x_{AB}
1	1	1
0	1	0
1	0	0
0	0	0

Table 2.2: Coding genetic interaction. Interaction between markers x_A and x_B is the pairwise logical 'and 'of their profiles.

2.3.7 *Phenotype prediction*

LMMs can be trained and used to predict phenotypes from genetic markers. Observations the model is trained on are referred to as *in-sample*, those for which predictions are made are referred to as *out-of-sample.* A combined matrix for the genetic markers is generated by stacking the in-sample, X, and out-of-sample, X_*, markers:

$$X^{all} = \begin{bmatrix} X \\ X_* \end{bmatrix} \tag{2.11}$$

R is computed as before (Equation 2.6) but is now referred to as R^{all}. Different sections of R^{all} correspond to covariance matrices of subsets of the data:

$$R^{all} = \begin{bmatrix} R & R_\times^T \\ R_\times & R_* \end{bmatrix} \tag{2.12}$$

R is the in-sample covariance matrix, R_* is the out-of-sample covariance matrix, and $R_\times$ is the cross covariance matrix between in-sample and out-of-sample markers. σ_g^2 and σ_n^2 are estimated having trained the model on the in-sample individuals. The joint distribution of the known, y, and unknown, y_*, phenotype vectors are modelled as:

$R_\times^T$ *is the transpose of* $R_\times$

$$\begin{bmatrix} y \\ y_* \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} \sigma_g^2 R + \sigma_n^2 & \sigma_g^2 R_\times^T \\ \sigma_g^2 R_\times & \sigma_g^2 R_* \end{bmatrix}\right) \tag{2.13}$$

which allows the prediction of the out-of-sample phenotype, y_*:

$$y_* = \sigma_g^2 R_\times^T (\sigma_g^2 R + \sigma_n^2)^{-1} y \tag{2.14}$$

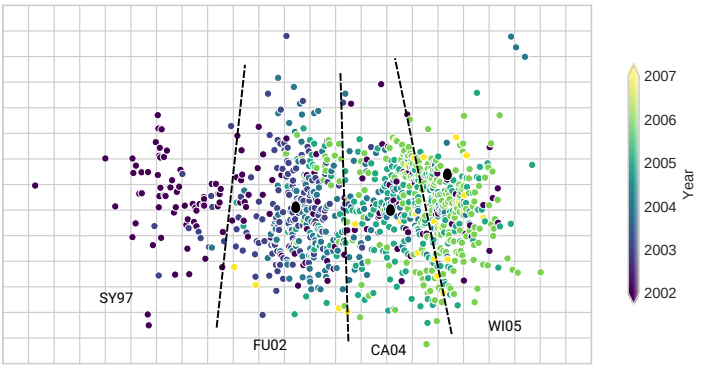
Equation 2.14 is known as the Best Linear Unbiased Predictor (BLUP). BLUP has been applied extensively to predict phenotypes in animal breeding (Robinson 1991). It also generalises to mvLMMs and can include fixed effects. For more details see Casale (2016).

2.4 PREDICTING ANTIGENIC COORDINATES OF INFLUENZA VIRUSES

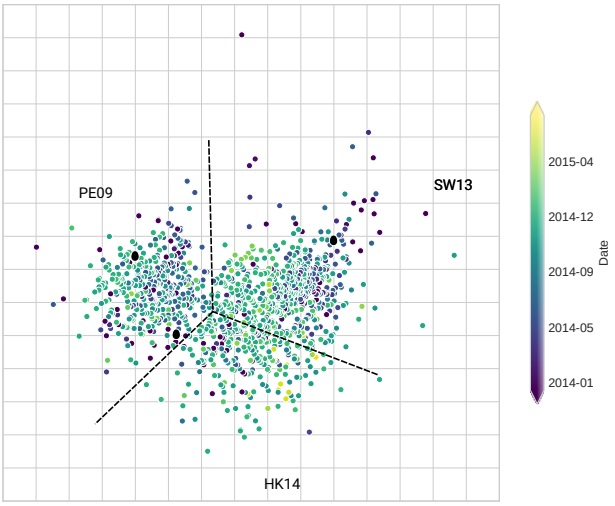
I tested the accuracy of BLUP for predicting antigenic coordinates of influenza viruses. I used three published human A(H3N2) influenza

The Russell et al. (2008b) and Li et al. (2016) maps look qualitatively different to the Smith et al. (2004) map. This is addressed in detail in Appendix A.

virus antigenic maps for which many antigens have an associated HA1 sequence: the Smith et al. (2004) map, which contains 273 antigens isolated between 1968 and 2003 (Figure 1.4); the Russell et al. (2008b) map, which contains approximately 13,000 antigens isolated between 2002 and 2007, 1,416 of which have HA1 sequences (Figure 2.4a); and the Li et al. (2016) map, which contains 1,226 antigens with HA1 sequences that were isolated between the beginning of 2014 and July 2015 (Figure 2.4b).



(a) Russell et al. (2008b).



(b) Li et al. (2016).

Figure 2.4: Human A(H3N2) influenza virus antigenic maps from Russell et al. (2008b) and Li et al. (2016). Black egg symbols indicate vaccine antigens; dashed lines delimit different antigenic variants referred to in the text.

There is a gap in the time these maps cover from 2007 to 2014 between the Russell et al. (2008b) and Li et al. (2016) maps and from 2015 to the present day since the Li et al. (2016) data was published. However, between 2007 and 2014, no major antigenic variants other than Wisconsin/67/2005 A(H3N2)-like (WI05) and Perth/16/2009 A(H3N2)-like (PE09) circulated, and these variants are both repre-

sented in these two maps. Similarly, since 2015, no major antigenic variants other than those that are present in the Li et al. (2016) map have circulated. Therefore, although there are temporal gaps, these three datasets cover all major antigenic variants of human seasonal A(H3N2) influenza viruses that have circulated since the A(H3N2) subtype began circulating in humans in 1968. Attempts to generate a single map containing A(H3N2) viruses from 1968 to the present day are ongoing, but have been hampered by changes in the types of red blood cells that agglutinate influenza viruses, and additional reasons outlined in Subsection 1.3.3.

2.4.1 *Measuring BLUP error*

I used a four-fold approach to measure BLUP error on each dataset. Each dataset was randomly partitioned into four as near to equal sized folds. Then, antigenic coordinates of one fold (out-of-sample) were computed having learnt the parameters in Equation 2.14 from the remaining three folds (in-sample). Koel et al. (2013) showed that the antigenicity of human A(H3N2) influenza viruses is determined by HA positions 109-301. Therefore, I restricted all analyses to using only HA positions 109-301. For more details see Section 2.3.

The error associated with a predicted antigenic coordinate can be expressed as the Euclidean distance between the predicted position and the true position. Distances, and therefore this error, cannot be negative, which may contribute to the positive skew of prediction error distributions (Figures 2.5–2.7). The median is a better measure of central tendency than the mean for skewed distributions, although both statistics are reported in Figures 2.5–2.7. The grand median of the errors for each dataset are Smith et al. (2004): 1.02 AU; Russell et al. (2008b): 1.54 AU; and Li et al. (2016): 1.03 AU. For comparison, Smith et al. (2004) predicted unknown HI titres from an antigenic map with an error of 0.83 AU.

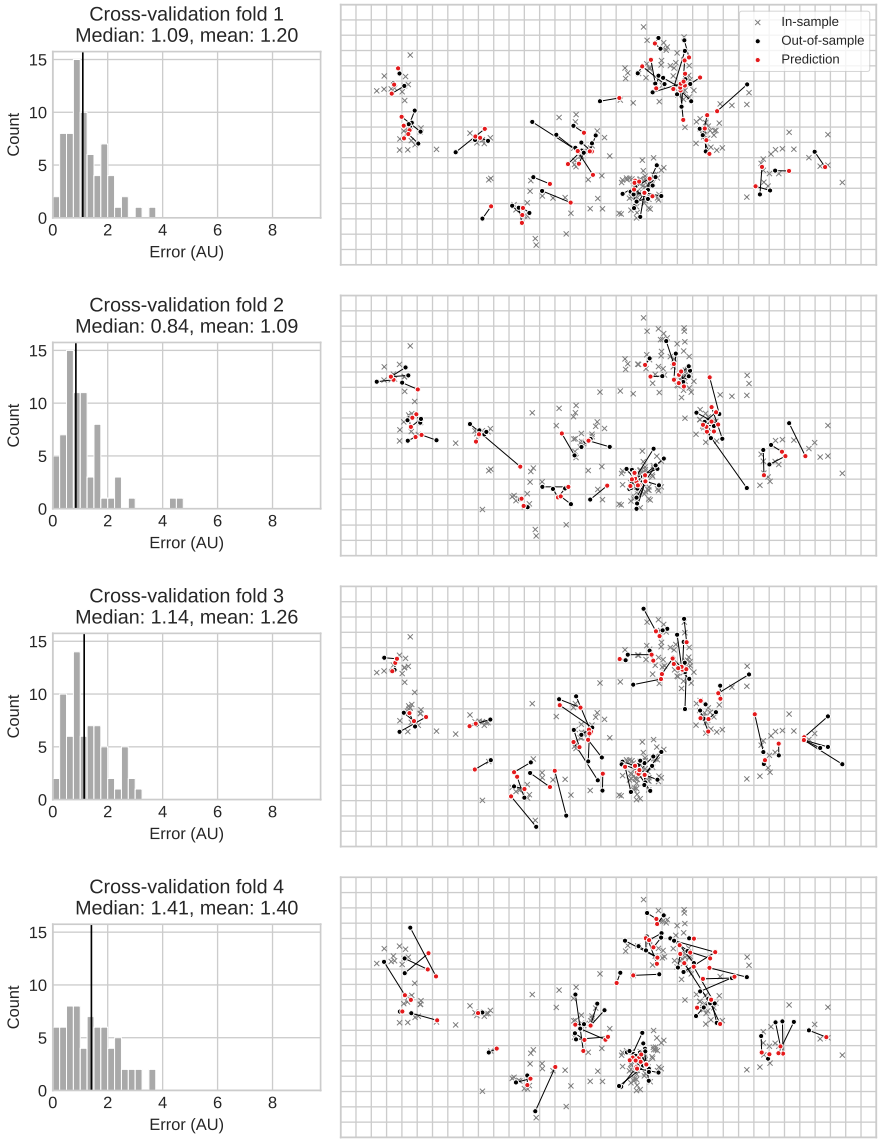


Figure 2.5: BLUP error estimated on the Smith et al. (2004) antigenic map. Each row shows of of the four random folds of the data; antigenic maps are shown on the right, histograms of the error distribution on the left. In the map, the predicted position of each out-of-sample antigen is connected to the true position by a line. Vertical lines on histograms indicate the median. This map is the same as in Figure 1.4 except it has been rotated 90° counter-clockwise and the antisera are not shown.

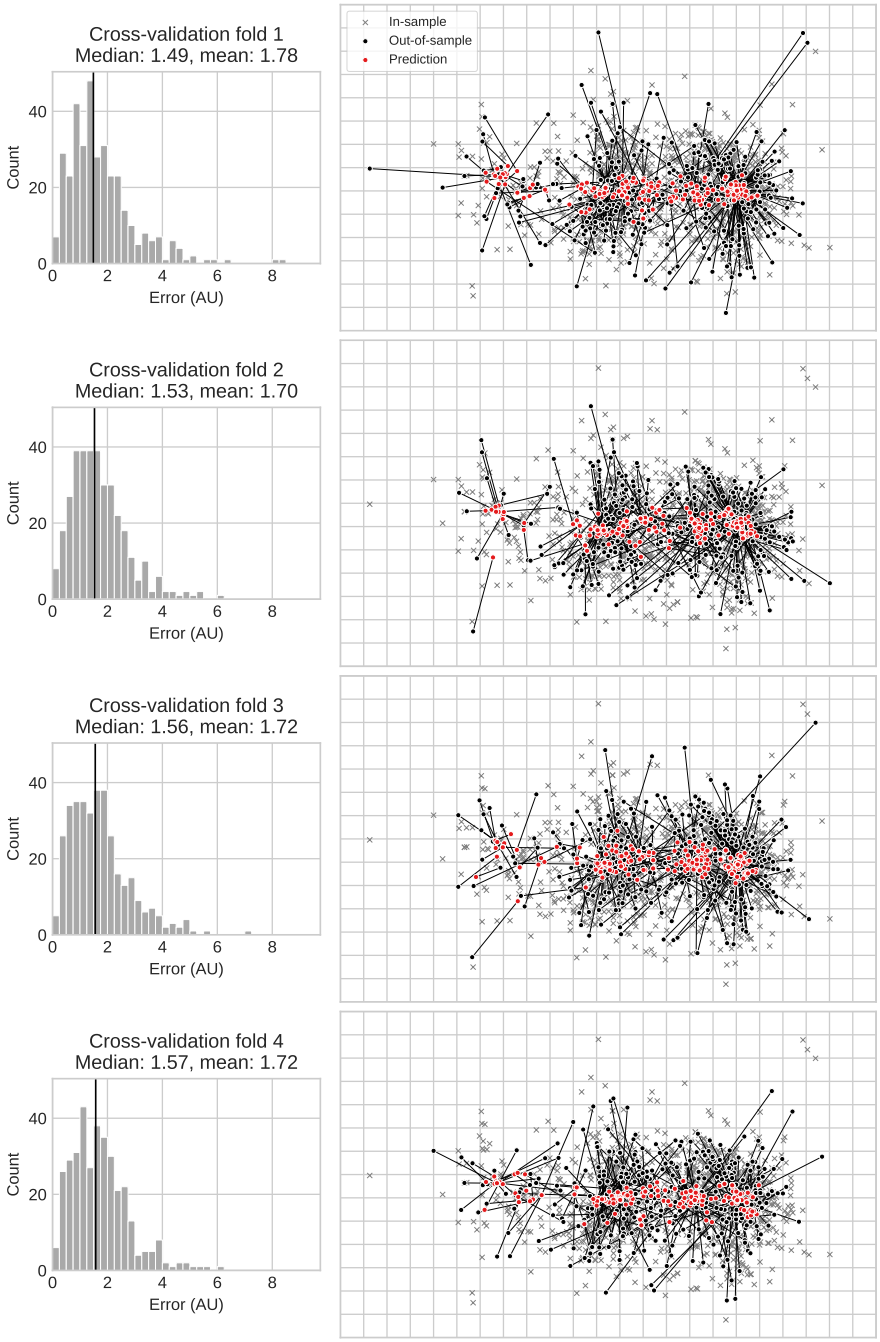


Figure 2.6: BLUP error estimated on the Russell et al. (2008b) antigenic map. See Figure 2.5 for a description.

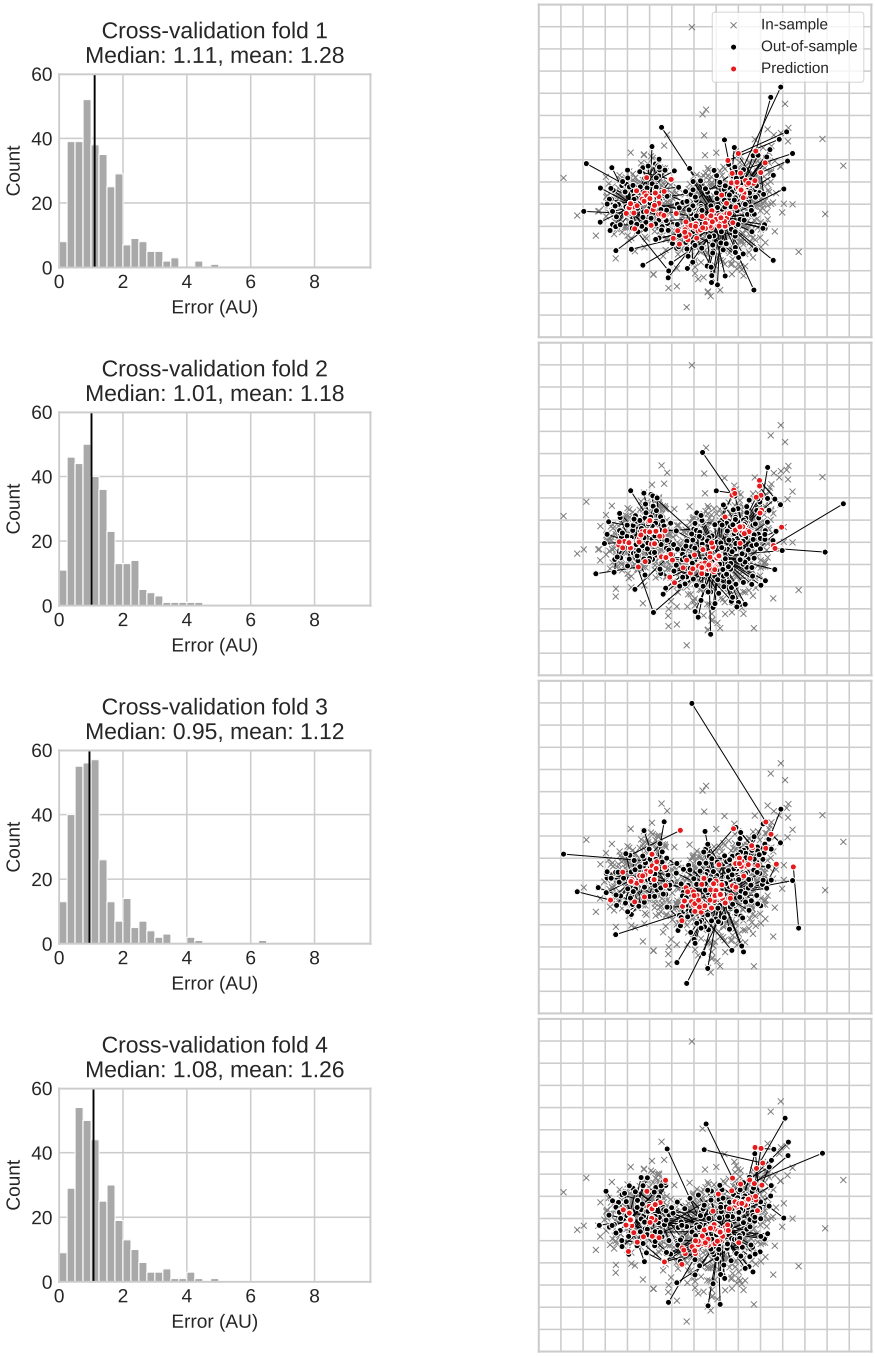


Figure 2.7: BLUP error estimated on the Li et al. (2016) antigenic map. See Figure 2.5 for a description.

2.4.2 *Discussion*

BLUP yields highly accurate predictions of antigenic phenotypes in the three datasets investigated here. This is true even for the Smith et al. (2004) dataset which has notably sparser temporal sampling than the other two datasets. This suggests there is sufficient signal in even relatively small datasets to predict antigenic phenotypes from HA sequences. Traditionally, surveillance centres have used an assay-first approach where all surveillance samples were assayed, and a subset of those were selected for genetic sequencing. Given that methods like BLUP can accurately predict antigenic phenotypes from HA sequence, the relatively low price and ease of sequencing makes a sequence-first approach appealing.

The current need to assay many isolates may be made superfluous by these methods. Research and surveillance effort may be better utilised generating smaller, high quality maps, that would provide quality training data for predicting antigenic phenotypes of all viruses that are sequenced.

Any method for predicting antigenic phenotypes using HA sequence may not accurately predict the phenotypes of viruses that are antigenically distinct from those the method has been trained on. An extreme example to illustrate this point would to predict the antigenic coordinates of a virus from the Hong Kong 1968 antigenic cluster having trained a model on the Li et al. (2016) antigenic map. The true antigenic position of the virus would be many tens of antigenic units away from the Li et al. (2016) map which the model could not predict. Therefore, a sequence-first surveillance approach must prioritise conducting titrations on viruses that exhibit genetic variation at and around sequence positions known to cause antigenic change, in order not to miss emerging antigenic variants. This is highly feasible given our detailed understanding of HA positions responsible for antigenic change in influenza viruses (Wiley et al. 1981; Wilson and Cox 1990; Lewis et al. 2011; Koel et al. 2013; Koel et al. 2014; Lewis et al. 2014; Abente et al. 2016).

Methods like BLUP may be able to extract components of antigenic variation that can be explained by genetic variation, thereby removing noise and easing interpretation. I have called this novel idea *map sanitising*. It is currently untested and tangential to the main subject of this thesis. I therefore introduce and discuss it more thoroughly in Appendix A.

BLUP predictions are used in this thesis to predict the antigenic coordinates of strains that were sequenced but not titrated in data presented in Chapter 3. This allowed me to substantially increase the number of viruses in those analyses.

3

QUANTIFYING THE RELATIONSHIP BETWEEN ANTIGENIC MISMATCH AND VACCINE EFFECTIVENESS

CHAPTER OUTLINE In this chapter I perform a quantitative comparison of VE and antigenic mismatch, using published VE estimates and antigenic mismatch measurements derived from antigenic maps. I then model the relationship to estimate the increase in VE that can be expected if antigenic mismatch in influenza vaccines were ameliorated.

3.1 INTRODUCTION

It is widely acknowledged that influenza is an antigenically variable pathogen, and that vaccine strains must be updated when antigenically novel viruses emerge and circulate widely. When the vaccine strain is antigenically dissimilar to circulating strains, the vaccine is said to be antigenically mismatched. Throughout this chapter I will refer to this phenomenon simply as *mismatch*.

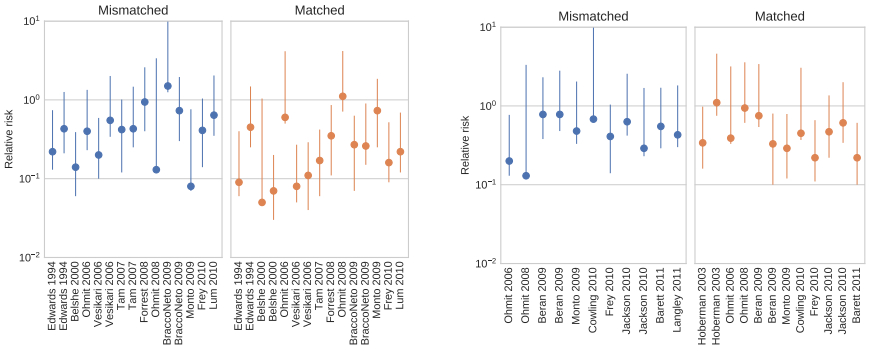
It is often stated that mismatch reduces VE (Paules et al. 2017; Lyons et al. 2018; Erbelding et al. 2018; Krammer 2019). There is strong theoretical reasoning to support this claim; hosts develop antibodies that are specific to strains previously encountered during infection or vaccination, therefore closer antigenic matches between prior exposure and a subsequent infection should yield higher affinity antibodies to the infecting strain.

3.1.1 *Existing data linking VE and mismatch*

However, data linking mismatch to VE is complex and sometimes contradictory. Some studies on individual influenza seasons do find the negative relationship between mismatch and VE that is predicted by the above mechanism (Zimmerman et al. 2016; Flannery et al. 2016). Findings from other studies are either ambiguous or would imply a positive relationship between mismatch and VE (Skowronski et al. 2007; Belongia et al. 2011; Fielding et al. 2016). Aggregating these data is difficult due to loose definitions of low or high VE and mismatch, and methodological differences between studies, and biases in VE estimates derived from studies that use the Test Negative Design (TND) (Jackson and Nelson 2013; De Serres et al. 2013; Sullivan et al. 2016; Foppa et al. 2013).

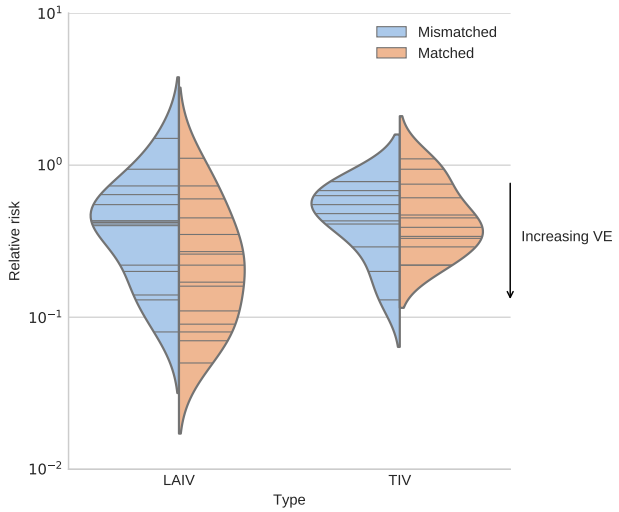
Studies that include multiple influenza seasons should exhibit lower methodological variation, (but different TND biases between seasons

may still exist). One study, spanning three influenza seasons, found large variation in VE, and that the highest VE coincided with the lowest proportion of mismatched strains (Belongia et al. 2009). A 2013 meta-analysis attempted to clarify the link between vaccine efficacy and mismatch (Tricco et al. 2013). The authors calculated separate Relative Risk (RR) scores for matched and mismatched influenza infections from published Randomly Controlled Trials (RCTs). Mismatch classification was based on data in the primary sources, or on HI assays using ferret antisera. The analysis showed a slight trend for increasing RR (decreasing VE) during mismatch seasons, but with substantial overlap between the distributions (Figure 3.1).



(a) LAIV

(b) TIV



(c) Summary

Figure 3.1: RR from the Tricco et al. (2013) meta-analysis. RR measured in RCTs, for (a) LAIVs and (b) Trivalent Inactivated Vaccines (TIVs). X-axis labels refer to studies cited by Tricco et al. (2013). (c) Summary of groups in (a) and (b) using Kernel Density Estimation (KDE). Individual measurements within each group are shown as horizontal lines inside the KDE.

Tricco et al. (2013) highlight that their treatment of mismatch as binary is overly simplistic and that "cross protection inferred by mismatch strains should be analyzed as a continuum in the future". There are two components of mismatch that are often used without acknowledging the difference.

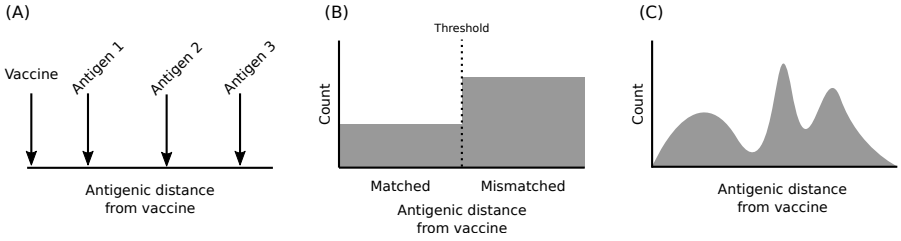


Figure 3.2: Components of antigenic mismatch. (A) Mismatch may relate to the degree of cross reactivity between an antigen and antiserum. Antigen 3 is more mismatched than antigen 1. (B) Mismatch may also relate to the number of circulating strains that are classified as either matched or mismatched. (C) This study uses 'mismatch distributions' which flexibly represent the abundance of circulating strains across the mismatch continuum.

The first component relates to the degree of reactivity between antiserum and antigen, which manifests itself as distance in an antigenic map (Figure 3.2A). For instance, in the 1997-1998 influenza season antigenically novel A/Sydney/1997 (H3N2)-like (SY97) viruses replaced A/Wuhan/1995 (H3N2)-like (WU95) viruses. Jong et al. (2000) showed that antisera from ferrets vaccinated with WU95 had a substantially lower HI titre to SY95. In an antigenic map, a WU95 antiserum would be further away from a SY97 antigen than a SY97 antiserum. More recently, Xie et al. (2015) show similar patterns for the transition from A/Texas/2012 (H3N2)-like viruses to A/Switzerland/2013 (H3N2)-like viruses. These studies stress the importance of the fold change in reactivity between antisera raised against the vaccine antigen and the antigens that circulated.

The second component is the proportion of circulating viruses that are mismatched (Figure 3.2B). In the above example, if only 1 % of circulation comprised the antigenically novel SY97-like viruses in the '97-'98 season, then the majority of strains would still have been antigenically matched, and VE would (presumably) have been higher. Numerous studies compute this by measuring the proportion of isolates that exceed a threshold HI titre to reference antisera (Skowronski et al. 2007; Belongia et al. 2009; Fielding et al. 2016; Skowronski et al. 2017). These approaches require selecting a threshold titre and classifying strains as either matched or mismatched.

To summarise, a season with 'low antigenic mismatch' may be one in which either the circulating strains are at a close antigenic distance to the vaccine, or one in which only very few mismatched strains circulated. In reality, both effects combine making a binary classifi-

cation system unsuitable. Here, I use positions of vaccine antigens and circulating strains estimated by antigenic cartography as the basis for assessing mismatch distance (Figure 3.2C). Antigenic coordinates vary continuously, therefore a threshold titre to define matched vs. mismatched is not required. Further, I use sampling strategies that allow variation in both the mismatch distance, and the proportion of strains that circulated at different mismatch distances to filter through to the fit of summary figures.

VE is typically measured by national bodies, so VE estimates relate to specific influenza seasons within a country. Country- and season-specific mismatch distributions were derived using knowledge of the vaccine strains used in by a country in a particular season, and isolation location and date of strains. These mismatch distributions were paired with the associated VE measure for that country and season. All analyses were conducted on the A(H3N2) influenza virus subtype. These data allowed me to conduct the first quantitative analysis on the relationship between mismatch and VE.

3.1.2 *Measuring VE*

This subsection briefly introduces how VE is estimated using the TND, and how the raw data counts from a TND study are used to calculate VE. The TND enables VE to be monitored across seasons using a *case-control strategy*. TND studies are cheaper, and whilst RCTs are better controlled, they only provide temporal-, geographic- and subtype-specific estimates of VE (Fukushima and Hirota 2017).

Individuals that present at medical institutions with ILI are tested for the presence of influenza virus. The gold-standard assay for this is Polymerase Chain Reaction (PCR) to detect the presence of influenza virus RNA in a clinical sample. Influenza positive individuals are matched to influenza negative individuals and vaccination status is ascertained by self reporting, or by reference to a database. This strategy controls for medical seeking attention between vaccinees and non-vaccinees, however the estimates may not extrapolate to non-medical seeking sections of the population (Belongia et al. 2009). Relative proportions of individuals that do and do not have influenza, and whether of not they were or were not vaccinated are used to estimate VE. Formally, a contingency table is constructed containing counts in each group and VE is computed from the Odds Ratio (OR) (Table 3.1, Equation 3.1).

$$\begin{aligned} \text{VE}\% &= 1 - \text{OR} \\ &= 1 - \frac{I_v/U_v}{I_u/U_u} \end{aligned} \tag{3.1}$$

	Infected	Uninfected
Vaccinated	I_v	U_v
Unvaccinated	I_u	U_u

Table 3.1: TND contingency table.

VE estimates are widely reported with a 95 % confidence interval which is derived from the standard error of the OR (Morris and Gardner 1988):

$$SE(\log OR) = \sqrt{\frac{1}{I_v} + \frac{1}{U_v} + \frac{1}{I_u} + \frac{1}{U_u}} \tag{3.2}$$

The $100(1-\alpha)\%$ confidence interval is calculated by first computing Y and Z:

$$\begin{aligned} Y &= \log OR - (N_{1-\alpha/2} \times SE(\log OR)) \\ Z &= \log OR + (N_{1-\alpha/2} \times SE(\log OR)) \end{aligned} \tag{3.3}$$

where $N_{1-\alpha/2}$ is the value from the standard normal distribution of the $1-\alpha/2$ percentile. The range of the confidence interval is then generated by exponentiating Y and Z.

3.2 METHODS

3.2.1 *Mismatch distributions*

Antigenic maps of data generated at the Centers for Disease Control and Prevention (CDC), U.S.A., and Victorian Infectious Diseases Reference Laboratory (VIDRL), Australia, as part of routine work conducted by the WHO GISRS (Russell et al. 2008a), were used as the source of antigenic data. For each season (Table 3.2), an antigenic map was taken from the subsequent WHO vaccine consultation meeting; i.e. for the 2016–2017 northern hemisphere season, a map from the meeting that occurred in September 2017 was used.

Virus strains with a sequenced HA, but that were not titrated for inclusion in an antigenic map had their antigenic coordinates predicted using BLUP (Chapter 2). BLUP can be used to accurately predict antigenic coordinates (Subsection 2.4.1), and allowed larger sample sizes to be used in this analysis (Figure 3.3 and Appendix C).

HA amino acid sequences were accessed from the Global Initiative on Sharing All Influenza Data (GISAID) database on 9th September 2018 (Shu and McCauley 2017) using the following query: type, A; H, 3; N, 2; host, human; submission from, 2000-01-01; submission to, 2018-09-09; required segments, HA. One empty sequence was removed: EPI ISL 285510. Three sequences labelled as influenza B viruses were also removed: EPI ISL 170670, EPI ISL 301415, EPI ISL 255542. Sequences

Country, Season	Mean	Median	Titrated	Predicted	Total	Figure
Australia, 2007	2.33	2.28	100	7	107	C.1
Australia, 2008	3.36	3.21	140	9	149	C.2
Australia, 2010	1.46	1.27	85	34	119	C.3
Australia, 2012	3.08	3.11	273	170	443	C.4
Canada, 2006–2007	1.26	1.08	16	29	45	C.5
Canada, 2007–2008	2.18	2.35	9	0	9	C.6
Canada, 2008–2009	3.47	3.26	17	11	28	C.7
Canada, 2010–2011	3.30	3.24	5	146	151	C.8
Canada, 2011–2012	2.75	2.81	4	72	76	C.9
Canada, 2012–2013	3.22	3.23	5	171	176	C.10
Canada, 2014–2015	3.72	4.13	14	493	507	C.11
Canada, 2016–2017	3.39	3.40	19	2109	2128	C.12
China, 2012–2013	3.17	3.23	11	43	54	C.13
New Zealand, 2013	2.22	2.36	4	17	21	C.14
South Africa, 2014	3.74	3.75	0	9	9	C.15
Spain, 2008–2009	2.83	2.90	0	212	212	C.16
Spain, 2011–2012	2.86	2.85	0	46	46	C.17
Spain, 2013–2014	3.34	3.31	0	248	248	C.18
Spain, 2014–2015	3.43	4.14	0	503	503	C.19
U.K., 2011–2012	2.74	2.82	1	41	42	C.20
U.K., 2012–2013	3.10	3.08	0	22	22	C.21
U.K., 2014–2015	3.35	4.14	0	392	392	C.22
U.S.A., 2007–2008	1.14	1.20	128	771	899	C.23
U.S.A., 2010–2011	3.12	3.17	699	475	1174	C.24
U.S.A., 2011–2012	3.04	2.95	1286	442	1728	C.25
U.S.A., 2012–2013	3.29	3.29	1362	1408	2770	C.26
U.S.A., 2014–2015	3.43	3.94	833	1965	2798	C.27
U.S.A., 2015–2016	2.43	2.39	422	968	1390	C.28
U.S.A., 2016–2017	3.54	3.45	820	2693	3513	C.29
U.S.A., 2017–2018	3.34	3.28	242	1370	1612	C.30

Table 3.2: Summary of antigenic mismatch measurements. Mean and Median are summary statistics of antigenic distances between strains that circulated in a country and season and the mean vaccine location. Titrated, Predicted and Total are counts of strains of each type in the map. Figure refers to the figure number showing the map and full distribution of distances in Appendix C.

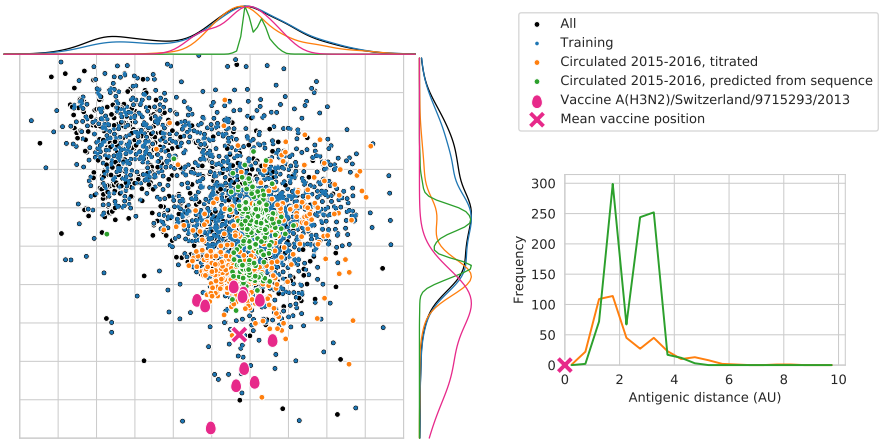
were aligned in MAFFT version 7.407 (Katoh and Standley 2013). The alignment was visually inspected and trimmed to positions 1–328 of HA1 (Burke and Smith 2014) in AliView (Larsson 2014). This alignment formed the sequence database from which antigenic coordinates of strains that were not titrated were predicted from. A LMM was trained on all antigens in a map that had a sequence in the database. Matching strains between the antigenic map and sequence database was done strictly whereby all fields in the virus identifier had to match. Then, strains that circulated in a time period that the map is being used for were predicted using BLUP (Chapter 2).

For each country-season combination I generated a 'mismatch distribution' which consists of the set of antigenic distances between each strain surveilled in a given country during a given season to that of the vaccine strain. Sometimes there are multiple different passages of the vaccine antigen and it was not possible to ascertain which specific passage was used by the vaccine manufacturer. In these cases, mismatch distributions were measured against the mean antigenic position of all vaccine antigens.

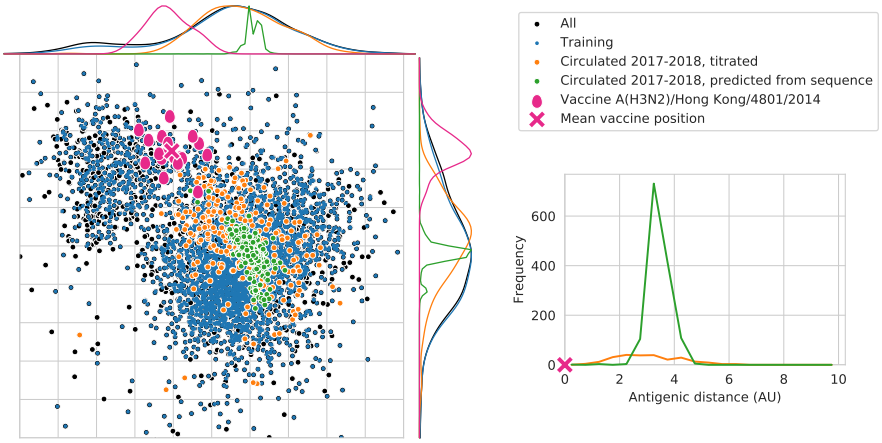
Sometimes multiple maps for a single season are available because the titrations originate from different laboratories, or a single laboratory conducted multiple assays using erythrocytes from different species. In seasons with multiple maps, all variant maps were used to generate independent mismatch distributions, which were then compared. A decision on which antigenic map to use was made based on the sample size and erythrocyte preference during that time period (see Subsection 1.3.3). See Appendix B for a comparison of mismatch distributions when multiple maps were available and an explanation of all decisions made. Antigenic maps used to derive mismatch distributions for the U.S.A. 2016–2017 and 2017–2018 influenza seasons are shown in Figure 3.3. See Appendix C for similar figures for all country-season combinations and Table 3.2 for a summary of all mismatch distributions.

3.2.2 *Sources of VE estimates*

I used VE estimates from a 2016 meta-analysis (Belongia et al. 2016) and additional VE estimates from subsequent studies (Table 3.3). The standard error of the OR, from which VE is derived, is normally distributed on the log-scale, therefore, all regression analyses were conducted on log-transformed OR values.



(a) U.S.A. 2015–2016



(b) U.S.A. 2016–2017

Figure 3.3: Antigenic maps used to derive mismatch distributions. Orange and green antigens circulated in the given season and country. Orange antigens were titrated, green antigens were sequenced and their antigenic coordinates were predicted using BLUP. Blue antigens were titrated and sequenced so comprise the BLUP training data. Vaccine antigens are pink, and the mean vaccine position is shown by a cross. All other antigens in the map, which circulated in a different season and/or in a different country are shown in black. Maps also display marginal densities of each dataset at the top and right. Mismatch distributions are shown in the right panel. The lines trace the height of histogram bins which have a width of 0.5 AU. Appendix C contains these figures for all country-season combinations.

Country, Season	1°	2°	VE	VE 95% CI
Australia, 2007	B-16	Belongia et al. (2016)	68	32–85
Australia, 2008	B-16	Belongia et al. (2016)	-66	-349–39
Australia, 2010	B-52	Belongia et al. (2016)	3	-495–84
Australia, 2012	B-52	Belongia et al. (2016)	46	21–63
Australia, 2012	B-46	Belongia et al. (2016)	35	-11–62
Australia, 2012	B-54	Belongia et al. (2016)	13	-20–36
Canada, 2006-2007	SPSN (2017)	-	41	6–63
Canada, 2007-2008	B-22	Belongia et al. (2016)	57	32–73
Canada, 2008-2009	SPSN (2017)	-	55	33–70
Canada, 2010-2011	B-42	Belongia et al. (2016)	39	14–57
Canada, 2011-2012	B-43	Belongia et al. (2016)	51	10–73
Canada, 2012-2013	B-41	Belongia et al. (2016)	41	17–59
Canada, 2014-2015	B-62	Belongia et al. (2016)	-8	-50–23
Canada, 2016-2017	SPSN (2017)	-	42	18–59
China, 2012-2013	B-57	Belongia et al. (2016)	43	-30–75
New Zealand, 2013	B-55	Belongia et al. (2016)	61	32–77
South Africa, 2014	B-60	Belongia et al. (2016)	-18	-172–48
Spain, 2008-2009	B-38	Belongia et al. (2016)	56	21–75
Spain, 2011-2012	Gherasim et al. (2017)	-	29	-11–55
Spain, 2013-2014	Gherasim et al. (2017)	-	-18	-104–31
Spain, 2014-2015	Gherasim et al. (2017)	-	-15	-101–34
U.K., 2011-2012	B-35	Belongia et al. (2016)	23	-10–47
U.K., 2012-2013	B-49	Belongia et al. (2016)	26	-4–48
U.K., 2014-2015	B-61	Belongia et al. (2016)	-2	-56–33
U.S.A., 2007-2008	Belongia et al. (2011)	CDC (2019)	41	24–53
U.S.A., 2010-2011	B-47	Belongia et al. (2016)	54	42–64
U.S.A., 2010-2011	B-8	Belongia et al. (2016)	48	1–73
U.S.A., 2011-2012	B-34	Belongia et al. (2016)	39	23–52
U.S.A., 2012-2013	B-53	Belongia et al. (2016)	39	29–47
U.S.A., 2014-2015	Zimmerman et al. (2016)	CDC (2019)	6	-5–17
U.S.A., 2015-2016	Jackson et al. (2017)	CDC (2019)	43	4–66
U.S.A., 2016-2017	CDC (2017)	CDC (2019)	34	—
U.S.A., 2017-2018	Flannery et al. (2017)	CDC (2019)	25	13–36

Table 3.3: Summary of VE data. '1°' refers to primary sources; those prefixed 'B-' refer to reference within Belongia et al. (2016). '2°' refers to secondary sources. Primary references from Belongia et al. (2016) are not duplicated here.

3.3 RESULTS

3.3.1 *VE vs. antigenic mismatch*

VE is independent of mismatch until approximately 3 AU, and then declines with increasing mismatch (Figure 3.4). Figure 3.4 is probably an overly simplistic representation of the data. Mismatch distributions are often irregular (Appendix C) so the horizontal error bars, whilst not symmetrical, are a poor summary of those distributions. Similarly, it is difficult to asses the error in VE particularly around a mismatch of 3 AU where there is a high amount of overlap.

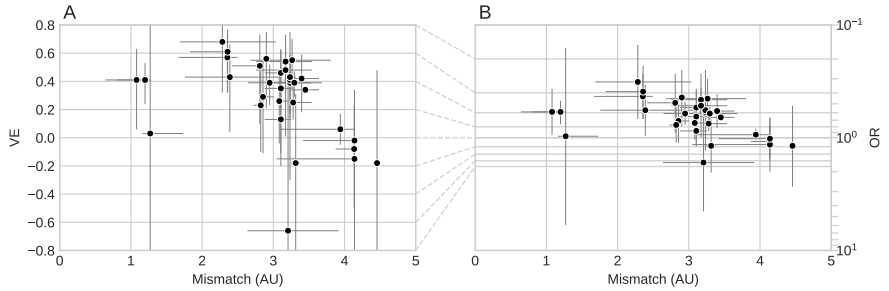


Figure 3.4: Relationship between VE, OR, and antigenic mismatch. Vertical bars show to the 95 % confidence interval of the VE estimate. Horizontal bars show the interquartile range of the mismatch distribution. (A) VE vs. antigenic mismatch. VE is expressed as a proportion of one. (B) OR vs. antigenic mismatch. A log-scale is used which highlights the symmetrical distribution of the y confidence interval.

3.3.2 *Sampling from VE and mismatch distributions*

In an attempt to better represent the data, for each country-season combination I generated a sample of 1000 pairs of VE and mismatch observations from their underlying distributions (Figure 3.5). For the y variable I sampled from the log(OR) distribution specified by the mean and 95 % confidence interval of the VE estimate. For the mismatch distributions, I modelled the vaccine and circulation antigenic coordinates using KDE. KDE is a non-parametric approach to estimate the probability distributions from which data are derived. Here, they flexibly capture the idiosyncratic distributions of antigenic coordinates of circulating and vaccine antigens, and therefore suitably represent confidence in the antigenic locations of viruses in a particular season. If very few strains are known for a particular season the KDE becomes broader (compare the KDEs for the nine strains known from Canada 2007–2008 to that of the 151 known for Canada 2010–2011, Figure 3.6). When a dataset contained more than five strains, KDE bandwidth was estimated using 3-fold cross validation. Otherwise, the bandwidth

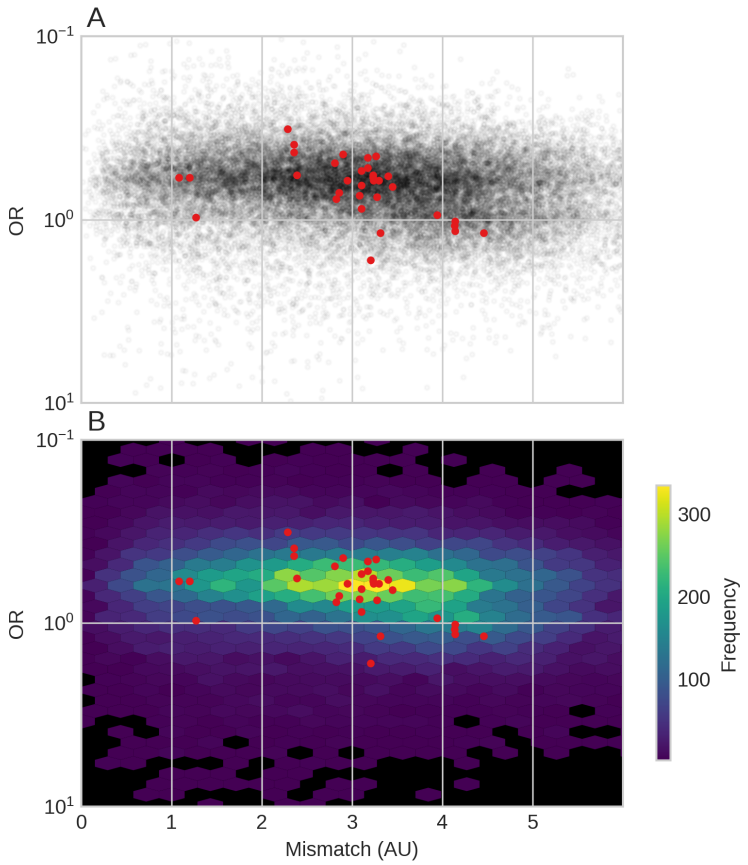
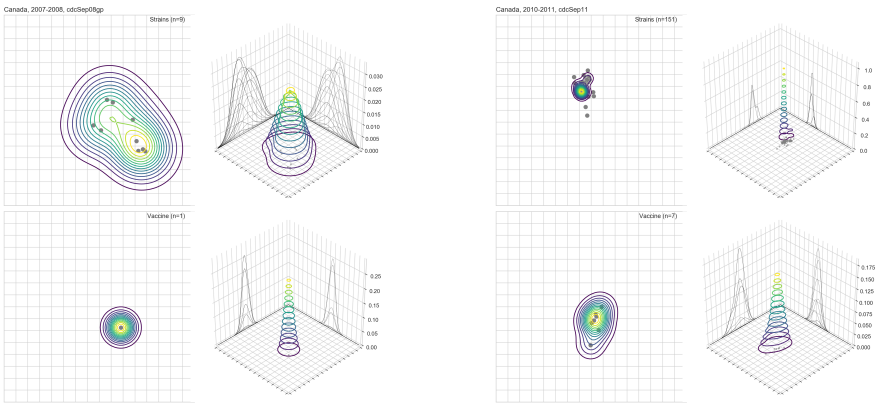


Figure 3.5: Joint distribution of VE and mismatch. (A) OR vs. antigenic mismatch having sampled the underlying distributions of data in both axes (see text for details). Red marks correspond to the point estimates that are visualised in Figure 3.4B. (B) Data in (A) plotted as a 2D histogram. The xy space is divided into hexagonal cells and the number of data points inside each cell is indicated using a colour scale (right). A cell with a count of zero is shown in black. Point estimates of x and y are shown in red.

was derived from estimates of map error based on the variance of antigenic coordinates of genetically identical strains (Figure A.1). KDEs for each country-season combination were sampled according to their probability density function, and used to generate 1000 pairs of vaccine and circulating strains. The antigenic distance within the pairs was computed and matched to a random sample from the log(OR) distribution (Figure 3.5).

Figure 3.5B shows that the distribution is tighter in y compared to x than is implied by Figure 3.4. The region between 1.2 AU to 2.2 AU which is poorly sampled when viewing only median mismatch, continues to show support for the approximately flat relationship between VE and mismatch at mismatch distances of less than 3 AU. Low VE outliers that are striking when looking at point estimates tend



(a) Canada 2007–2008 (b) Canada 2010–2011

Figure 3.6: KDEs used to represent antigenic coordinates of circulating strains and vaccine strains. (a) The top panel shows strains that circulated in Canada in the 2007–2008 influenza season. The bottom panel shows vaccine antigens used in this country and season. Antigenic coordinates of individual strains are shown with grey dots, contours display the probability density function of the KDE of each dataset. Left sub-figures show a top down view of the KDE. Right sub-figures show a 3D projection with the shape of the KDE also projected onto the margins. (b) The Canada 2010–2011 season data plotted as in (a).

to have wide x and y distributions, meaning they do not contribute strongly to the signal in the data.

3.4 MODELLING THE RELATIONSHIP

A predictive model of influenza virus VE should include some measure relating to mismatch (Figure 3.5). Additional factors such as subtype, country and vaccine type may explain additional variance in the data. Dissecting how different factors contribute to VE is clearly interesting scientifically, and useful for optimising vaccine strain selection. Parametrising these factors may inform the best areas to focus research effort and funding. *In lieu* of a wider and more complete dataset, I present here the first step in developing a more complete model. In all subsequent analyses in this chapter, log(OR), y is the response variable, and median antigenic mismatch, x, is the sole predictor.

3.4.1 *Weighting*

Confidence intervals on VE estimates are highly variable (Table 3.3). Ideally, tighter confidence estimates should have a greater influence on model fit. Conceptually, the likelihood of the model should increase if there is less residual error for a high confidence observation, than for

a low confidence observation. Given a predictor, $\mathbf{x}$, and a response, $\mathbf{y}$, a standard, non-weighted, linear model can be cast as:

$$\begin{aligned} y_i &= \alpha + \beta x_i + \epsilon_i \\ \boldsymbol{\epsilon} &\sim \mathcal{N}(0, \sigma^2) \end{aligned} \tag{3.4}$$

Where α is the effect size, β is the intercept, and error, ϵ follows a normal distribution with a mean of zero, and variance σ^2. This is known as Ordinary Least Squares (OLS) regression. In contrast, Weighted Least Squares (WLS) regression incorporates the variance of each observation by modelling the error with a multivariate normal distribution:

$$\begin{aligned} y_i &= \alpha + \beta x_i + \epsilon_i \\ \boldsymbol{\epsilon} &\sim \mathcal{N}(0, \boldsymbol{\Sigma}) \\ \boldsymbol{\Sigma} &= \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{bmatrix} \end{aligned} \tag{3.5}$$

The covariance matrix of the error distribution is filled with zeros, except for the leading diagonal which contains the variance of each observation, σ_i^2.

I conducted OLS and WLS using median antigenic mismatch as the predictor and log(OR) as the response, and found that model fit is insensitive to weighting (Figure 3.7). An explanation for this is that although observations are heteroscedastic, no regions of the data have particularly low variance to pull the WLS fit away from the OLS fit. Therefore, although variance is not uniform it appears to be randomly distributed throughout the dataset. Given the similarity between the OLS and WLS fits, all subsequent analyses were conducted using unweighted methods.

3.4.2 *Alternative models*

The relationship between mismatch and VE is clearly not linear (Figure 3.4). The bulk of the data between mismatch values of 2.5 AU to 3.5 AU fit well, but the data do not increase at low mismatch, nor decrease at high mismatch as the model predicts. I conducted a model comparison analysis to determine a more appropriate fit. Against the linear model (Equation 3.4, Figure 3.8a) I compared three additional models:

1. Offset linear:

$$\begin{aligned} y_i &\begin{cases} \alpha + \epsilon_i & x_i \leqslant \gamma \\ \alpha + \beta(x_i - \gamma) + \epsilon_i & x_i > \gamma \end{cases} \\ \boldsymbol{\epsilon} &\sim \mathcal{N}(0, \sigma^2) \end{aligned} \tag{3.6}$$

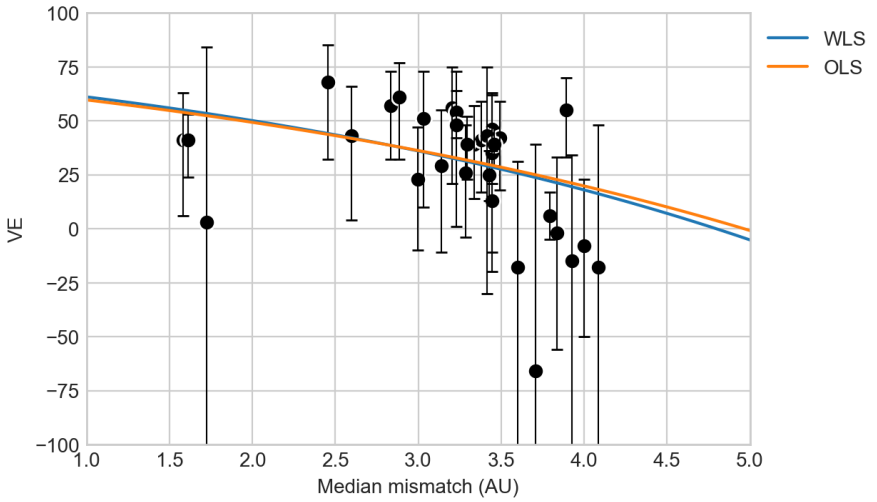


Figure 3.7: OLS and WLS fits to the VE vs. mismatch data. Vertical error bars show the 95 % confidence interval in the VE estimate. The bottom of some error bars are trimmed, see Table 3.3 and Figure 3.5 for untrimmed error bars. This figure is plotted in VE space, but regressions are conducted on log(OR) which causes the non-linear relationship in this figure. Figure 3.4 shows the mapping of VE to log(OR).

where y is a constant, α, up to a particular x value, γ, whereafter it becomes linearly dependent on x (Figure 3.8b). This was motivated *a posteriori* as a simple formulation with constant and linear phases.

2. Inverse logit:

$$\begin{aligned} y_i &= \gamma - \frac{\delta}{1 + e^{\beta(x_i - \alpha)}} + \epsilon_i \\ \boldsymbol{\epsilon} &\sim \mathcal{N}(0, \sigma^2) \end{aligned} \tag{3.7}$$

Here, y has two constant phases at low and high x which are linked sigmoidally (Figure 3.8c). The inverse logit model has been used to model the relationship between pre-vaccination HI titre and risk of influenza infection (Coudeville et al. 2010).

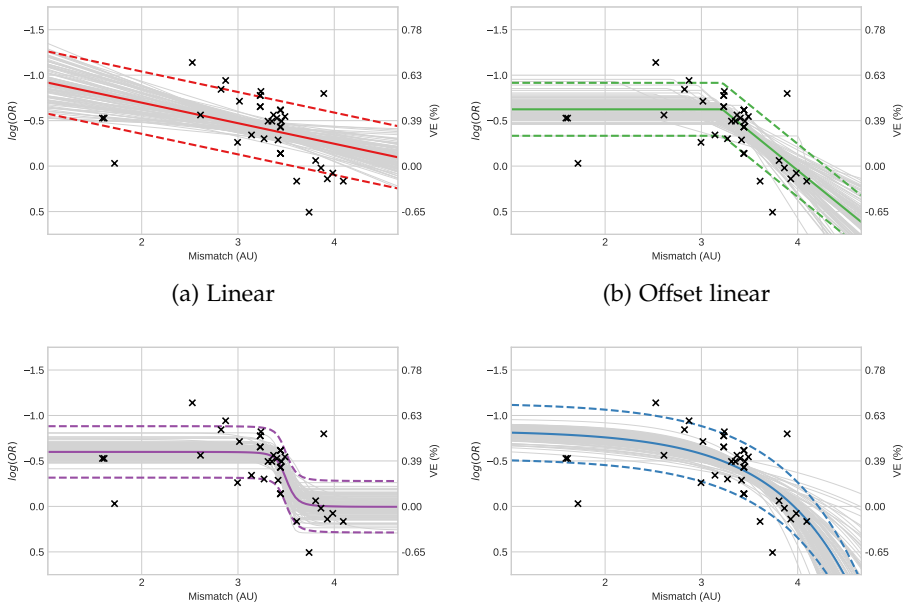
3. Exponential:

$$\begin{aligned} y_i &= \alpha + e^{\beta x_i} + \epsilon_i \\ \boldsymbol{\epsilon} &\sim \mathcal{N}(0, \sigma^2) \end{aligned} \tag{3.8}$$

One difference to the inverse logit model used by Coudeville et al. (2010) is that the authors bounded the inverse logit function such that it did not exhibit an inflection point. The exponential model here exhibits a similar shape (Figure 3.8d).

I used a Bayesian framework to sample posterior distributions of model parameters and used weak uninformative priors. A normal

distribution with mean zero and standard deviation of ten was used as the prior for parameters that can be any real number. A Half Cauchy distribution with $\beta = 10$ was used for parameters only requiring support for values greater than or equal to zero. Model parameters were sampled one million times using the Sequential Monte Carlo step algorithm implemented in PyMC3 (Salvatier et al. 2016). Parameter samples were assessed for convergence, and are summarised in Figure 3.8.



(a) Linear

(b) Offset linear

(c) Inverse logit

(d) Exponential

Figure 3.8: Model comparison of log(OR) as a function of mismatch. A random 100 of the one million posterior distribution samples are shown in light grey. The mean of each parameter was used to visualise the average model, overlaid as a solid coloured line. Dashed lines indicate plus and minus the mean of the sampled standard deviation of error from the mean model.

Poor suitability of linear model is shown by a divergence between the posterior samples and the data above 4 AU and below 2 AU mismatch. Under the linear model, samples are drawn such that they go through the bulk of the data at approximately 3.5 AU. This leads to poor extrapolation at the mismatch extremes, and despite the slopes being highly diverse, the linear model samples do not cover the extreme mismatch values well. The three other models suffer less from this issue, and have tighter distributions of samples at the extremes of mismatch.

The y-intercept of the models can be interpreted as the VE given a perfectly matched vaccine. y-intercepts of the offset linear and inverse logit models are relatively consistent with each other (means of -0.61 and -0.56 respectively), in contrast to the exponential model (mean of

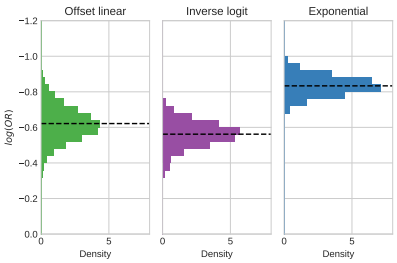


Figure 3.9: Posterior distribution of y-intercepts for offset linear, inverse logit and exponential models. The black dashed line indicates the mean.

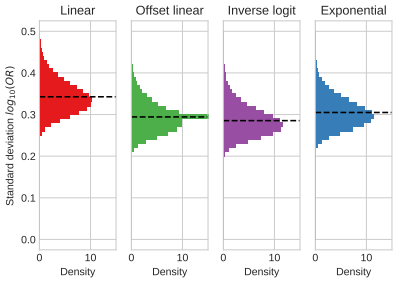


Figure 3.10: Posterior distribution of error parameters from VE-mismatch models. Each figure shows the distribution of σ for a different model. The black dashed line indicates the mean.

-0.83) (Figure 3.9). Precise y-intercept estimates should not be taken from this analysis due the paucity of data at low antigenic mismatch.

The VE predicted when mismatch tends to large values also has a clear interpretation, in that it should correspond to a VE of a vaccine with no cross reactivity to circulating strains. A vaccine that provides no cross reactivity with circulating influenza strains should produce $VE = 0$ or $OR = 1$. Conceptually, this could be a vaccine from a different influenza virus subtype, or even an entirely different pathogen. Ideally then, as mismatch increases, VE should tend to zero, and OR should tend to one. The only model investigated here that tends to a constant value at high mismatch is the inverse logit. Furthermore, it is encouraging that the constant value the inverse logit model tends to at high mismatch is $OR = 1$. Predictions from the other three models of VE given antigenic mismatch should not be extrapolated at mismatch distances above the range of the data. The offset linear model could be further extended such that it has a second constant phase at high mismatch.

The standard deviation of the error term in the models, σ, captures how much variance is not explained by mismatch in each model. The offset linear, inverse logit and exponential models all explain a similar amount of variance in VE as each other, and each explain a higher amount than the linear model (Figure 3.10). Similar estimates of this parameter, excluding the linear model, suggests other factors may be

		A			
		Linear	Offset linear	Inverse logit	Exponential
B	Linear	-	1.115	1.730	34.589
	Offset linear	0.897	-	1.551	31.012
	Inverse logit	0.578	0.645	-	19.996
	Exponential	0.029	0.032	0.050	-

Table 3.4: VE-mismatch Bayes factors. Bayes factors are presented as the marginal likelihood of model A divided by the marginal likelihood of model B.

important in explaining variance in VE, rather than requiring a more complex model that only incorporates mismatch.

3.4.3 *Quantitative model comparison*

Use of the Sequential Monte Carlo step algorithm enabled computation of marginal likelihoods, which can be used for model comparison. Ratios of marginal likelihoods between models (Bayes factors) quantify the relative support for one model over another, whilst compensating for model complexity.

Bayes factors can be approximated by the Bayesian Information Criterion (BIC), which is similar to the Akaike Information Criterion (AIC).

Marginal likelihoods for each model were as follows: linear, 8.435E-28; offset linear, 9.408E-28; inverse logit, 1.459E-27; exponential, 2.918E-26. A widely used heuristic is to declare 'substantial support' for the model in the numerator of the Bayes factor if the Bayes factor is greater than three (Kruschke 2014). All Bayes factors of all combinations of the linear, offset linear and inverse logit are in between a value of one and two (Table 3.4), suggesting there is no strong evidence to prefer any one of these three models over another. The exponential:linear and exponential:offset linear Bayes factors both exceed 30, and exponential:inverse logit is just under 20, which suggests a strong preference for the exponential model.

Equivocal Bayes factors among the non-exponential models, despite the inverse logit and offset linear models fitting the data better, and having lower σ parameter estimates, is due to their higher model complexity. The inverse logit and offset linear models require one and two additional parameters relative to the linear model. In a Bayesian framework, additional parameters penalise the marginal likelihood of a model by having diluted the prior probability of parameter values that are consistent with the data. The exponential model is preferred because it fits the data approximately as well as the offset linear and inverse logit models, but also has the lowest number of parameters (equal to that of the linear model).

3.4.4 *Qualitative model comparison*

The inverse-logit model is attractive despite having lower Bayes-factor support than the exponential model. As mismatch increases beyond the range of the data, the inverse-logit model is the only model that matches theoretical understanding and predicts VE tending to zero. Qualitatively, the inverse-logit and exponential models seem to fit the data between 2 AU to 4 AU equally well. At low levels of antigenic mismatch (below 2 AU) these two models diverge. Knowing which model best captures the underlying process at low antigenic mismatch is important because it enables prediction of the highest level of mismatch that may still provide optimal VE. In this region of the data the inverse-logit model does have lower residual error than the exponential model, so should be slightly preferred. However, there are only three data points with low antigenic mismatch, one of which has a particularly wide confidence interval, meaning that there is not a strong basis to prefer the inverse-logit over the exponential model.

3.5 DISCUSSION

In this chapter I have presented the first empirical data that provides a quantitative link between antigenic mismatch and VE. This framework has applications beyond refining our understanding of this long known relationship. Influenza virus vaccine strain selection attempts to select vaccine strains that are antigenically close to circulating viruses. The majority of vaccine antigen is currently produced in embryonated chicken eggs. Before the vaccine strain selection procedure, vaccine manufacturers identify candidate vaccine strains that grow well in eggs which are referred to as high-growth reassortant viruses. The vaccine strain choice is not among all strains that have been surveilled and antigenically characterised by the GISRS, but among the smaller subset of high-growth reassortants generated by vaccine manufacturers. Passaging influenza viruses in eggs can select for mutations in HA that improve virus attachment to chicken cells. Due to the proximity of positions that cause major antigenic change to the HA RBS (Koel et al. 2013), such mutations may also alter antigenic properties (Zost et al. 2017). Low VE has been attributed to antigenic mismatch derived from egg passaging (Skowronski et al. 2014; Chen et al. 2018). The quantitative framework presented in this chapter provides an opportunity to test the hypothesis that VE depends more on the antigenic position of the egg than the cell isolate. Concretely, for seasons where the antigenic positions of cell isolated and egg isolated vaccine strains are known, separate mismatch distributions could be derived, and which (if either) correlates better with VE could be tested.

Surveillance of human influenza isolates may be biased towards more extreme cases. Infections yielding more severe symptoms may be

more likely to cause individuals to seek medical care, and once seen by a medical professional, may be more likely to have virus isolated and submitted to surveillance programs. Antigenically novel influenza viruses may produce more severe infections, and therefore may be overly represented in surveillance. This may mean that mismatch distributions estimated in this study are biased towards higher values. This potential bias may contribute to the relative paucity of seasons with low mismatch estimates (Table 3.2).

This framework could also be used to test the hypothesis that a sanitised map (Appendix A) better represents true antigenic phenotypes than a raw antigenic map. If map sanitising removes noise, then sanitised maps should produce mismatch distributions with less variance, but that still recover a similar relationship with VE.

Additional variables of interest should also be included in further modelling that may either explain more variance in the data, or would have interesting parameter estimates for vaccination strategies. VE estimates are now routinely made for the main influenza types and subtypes that circulate in humans: the Victoria and Yamagata type B influenza virus lineages as well as the type A, H3N2 and H1N1 subtypes (Belongia et al. 2016). This study was conducted on the A(H3N2) subtype because it has shown greater antigenic variation in the past decade than influenza B viruses or the A(H1N1) subtype. Lower levels of antigenic variation in other types and subtypes may reduce the mismatch signal in VE for these data. Including data from influenza virus types and subtypes may enable more definitive estimation of precise quantitative relationship between mismatch and VE, and may provide measurements of type- and subtype-specific VE that are controlled for antigenic mismatch.

VE estimates and mismatch distributions were both derived from country specific data, which may lead to country-wise effects. For instance, there may be known or unknown differences between the application of TND in different countries. Likewise, there are large differences in influenza virus surveillance intensity from country to country, and differences in antigenic characterisation procedures in different WHO collaborating centres.

In addition, recent studies have theorised and shown that VE may wane through the course of an influenza season (Ray et al. 2019b), and have demonstrated methods to correct for it (Ray et al. 2019a). Therefore, time at which VE was estimated during a season should be included as an additional covariate in future modelling studies.

This chapter highlights the importance of antigenic match to influenza vaccines in a novel quantitative fashion. This framework should be used as a basis for elucidating variables that explain variance in VE estimates. In Chapter 5 I present one component of a framework to select antigenically advanced influenza vaccine strains and discuss

how this framework could ameliorate mismatch issues caused by the current WHO influenza vaccine strain selection process.

4

THE MOLECULAR BASIS OF ANTIGENIC CHANGE

CHAPTER OUTLINE A detailed understanding of molecular changes responsible for antigenic change in influenza viruses is fundamental to investigate what governs the antigenic evolution of these viruses, and any attempt to predict their antigenic evolution. In this chapter I review current knowledge of the molecular basis of antigenic change which has been derived from wet lab experiments that measure the antigenic impact of introducing substitutions into virus strains. These techniques have been applied to human A(H3N2) influenza viruses as well as other human viruses (A(H1N1), B(Victoria), B(Yamagata)), and non-human viruses (equine A(H3N8), swine A(H3N2), avian A(H5N1)).

For human A(H3N2) viruses, these wet lab analyses have only been conducted on strains that circulated between 1968 and 2003, so we do not know definitively the substitutions responsible for antigenic change in viruses that circulated since 2003. Therefore, in this chapter, I also present analyses to elucidate the amino acid substitutions responsible for antigenic change from 2003 until the present day. These analyses use published antigenic maps derived from GISRS data and consist of systematically searching for natural experiments that test the impact of substitutions, as well as applying association tests derived from a LMM. Substitutions reviewed and identified in this chapter are used as training data for the cluster transition substitution ranking framework presented in Chapter 5.

4.1 PREVIOUS WORK

4.1.1 *Human A(H3N2) 1968–2003*

Koel et al. (2013) determined the substitutions responsible for altering the antigenic phenotype of viruses from one cluster to another, and termed them *cluster transition substitutions*. To do so, Koel et al. (2013) initially sought to reduce the number of HA positions under consideration by transplanting large sections of HA1 between viruses to test which regions determined antigenic phenotypes. These experiments conclusively showed that antigenic phenotypes are determined by amino acid positions 109–301 of HA1. Within this reduced search space, Koel et al. (2013) then identified any positions where the consensus sequences of viruses in two adjacent antigenic clusters had different amino acids. They termed these substitutions *cluster difference substitutions*. Importantly, not all cluster difference substitutions may

be responsible for antigenic change. For instance, a substitution may evolve due to another selective pressure, such to alter HA binding, or due to genetic drift. The problem becomes identifying the cluster transition substitutions among the cluster difference substitutions.

To do this, Koel et al. (2013) inserted each cluster difference substitution into consensus strains from each cluster. Antigenic phenotypes of mutant viruses were measured by HI assays and analysed using antigenic cartography. In three out of ten cluster transitions, no single cluster difference substitutions was able to cause the transition, so combinations of multiple cluster difference substitutions were tested. Cluster transition substitutions in human seasonal A(H3N2) influenza viruses that were identified by Koel et al. (2014) are listed in Table 4.1.

Koel et al. (2013) also identified two accessory substitutions, which adjusted the direction of the antigenic change towards the subsequent cluster but did not substantially alter the antigenic distance.

Substitution(s)	Cluster transition
T155Y	HK68–EN72
Q189K	EN72–VI75
D193N, G158E	VI75–TX77
K156E	TX77–BK79
K189R, S159Y, Y155H	BK79–SI87
N145K	SI87–BE89
E156K	SI87–BE92
N145K	BE92–WU95
E158K, K156Q	WU95–SY97
Q156H	SY97–FU02

Table 4.1: Cluster transition substitutions in human seasonal A(H3N2) influenza viruses isolated between 1968 and 2003 (Koel et al. 2013).

4.1.2 *Human A(H1N1), human B(Victoria) and human B(Yamagata)*

Koel et al. (2013) also identified cluster transition substitutions in human seasonal A(H1N1), B(Victoria) and B(Yamagata) influenza viruses using the same methods described above. They are summarised in Table 4.2.

4.1.3 *Swine A(H3N2)*

Lewis et al. (2014) assessed the antigenic diversity of influenza A(H3N2) viruses isolated in swine populations in the U.S.A. using antigenic cartography. Due to the low number of isolates analysed it was not possible to unambiguously identify substitutions responsible for antigenic changes between clusters and outlier variants, except for N145K. Abente et al. (2016) tested the antigenic impact of substitutions iden-

tified by Lewis et al. (2014), as well as other substitutions, by introducing them into a prototypic virus from the Cyan antigenic cluster. The molecular determinants of three cluster transitions were identified (see Figures 1A, 1B, 1D, 1E (Abente et al. 2016), and Table 4.2).

4.1.4 *Equine A(H3N8)*

Lewis et al. (2011) conducted HI assays and antigenic cartography on equine A(H3N8) viruses isolated between 1968 and 2007. Three antigenic clusters were identified, and their molecular basis was elucidated (Table 4.2). 189K distinguishes viruses in the green antigenic cluster from viruses in the blue antigenic cluster, which have 189N, 189D, 189Q, or 189E (Lewis et al. 2011). The relative proportions of viruses in the blue antigenic cluster with those amino acids at position 189 in the Lewis et al. (2011) alignment are 15.3%, 1.4%, 77.8% and 5.6%, respectively. I conclude the substitution responsible for the blue to green transition as Q189K.

Colour names were used to identify antigenic variants by Lewis et al. (2011).

4.1.5 *Avian A(H5N1) clade 2.1*

Koel et al. (2014) elucidated substitutions responsible for antigenic variation in avian A(H5N1) clade 2.1 influenza viruses using reverse genetics. D183N and R189M induced the antigenic change from A/Indonesia/5/05 to A/Chicken/East Java/121/10 (Figure 3B in Koel et al. (2014)). These two substitutions, combined with I151T and Δ129, also caused the antigenic change from A/Indonesia/5/05 to A/Chicken/West Java/119/10 (see Figure 3C in Koel et al. (2014)). A/Chicken/West Java/119/10 is at similar angle from A/Indonesia/5/05 as A/Chicken/East Java/121/10, but is at a greater antigenic distance. We interpret D183N and R189M as having the same antigenic impact in both cases, which is modified by I151T and Δ129 to produce the A/Chicken/West Java/119/10 phenotype. Thus, D183N and R189M are both included once in the dataset with respect to these phenotype transitions. These substitutions are summarised in Table 4.2.

ΔN is notation for an amino acid deletion at site N.

4.2 METHODS

In the previous section I reviewed substitutions responsible for antigenic change in a variety of seasonal influenza viruses that are derived from wet lab experiments. Here I introduce computational approaches to address the same question.

I apply these approaches to human seasonal A(H3N2) influenza viruses, for which wet lab derived cluster transition substitution data end at 2003 (Koel et al. 2013). The computational approaches I present are strengthened by the large number of viruses now routinely se-

Host	Type	Subtype/Lineage	Substitution	Source
Avian	A	H5N1	A185E D183N I151T R189K R189M S129L S133A	Koel et al. (2014)
Equine	A	H3N8	E189K N159S	Lewis et al. (2011)
Human	A B	H1N1 Victoria Yamagata	K144E N159K N160Y	Koel et al. (2013)
Swine	A	H3N2	H155Y R189K N145K R189E N156K	Abente et al. (2016)

Table 4.2: Cluster transition substitutions in non-'human A(H3N2)' influenza viruses. Subtype-specific numbering is used throughout (Burke and Smith 2014).

quenced and antigenically characterised by the GISRS. There is substantial GISRS data from 2003 onwards which corresponds to when the wet lab derived cluster transition substitution data end. Antigenic maps using this data have been published (Russell et al. 2008b; Li et al. 2016) and the datasets were described in Section 2.4. I use the same two antigenic maps in this analysis.

I follow the same approach as Koel et al. (2013) to identify substitutions responsible for cluster transitions. I identify cluster difference substitutions and then test which among them are responsible for the antigenic change.

I use the term *cluster* here in a loose sense. Antigens in the Russell et al. (2008b) and Li et al. (2016) maps are not as distinctly clustered as in the Smith et al. (2004) map (Figure A.3). Therefore, the k-means clustering algorithm applied to the Smith et al. (2004) data to formally define clusters would not identify biologically meaningful units if it were applied to the Russell et al. (2008b) and Li et al. (2016) maps.

Differences between the Smith et al. (2004) map and the Russell et al. (2008b) and Li et al. (2016) maps, and potential reasons for it, are discussed in Appendix A.

Nevertheless, it is useful to subdivide these maps into regions, and there is an underlying biological basis to do so. Large regions in the

Russell et al. (2008b) and Li et al. (2016) maps have strains with low HA1 diversity. These regions are marked in Figure 2.4. Finding the substitutions responsible for antigenic change across an entire map can be broken down into testing what substitutions cause a change in the antigenic phenotype from one of these regions to the next. Throughout this chapter I use the term cluster to refer to these regions.

In all analyses, I use positions 109–301 in the HA globular head. This is the large section of HA1 which Koel et al. (2013) demonstrated is responsible for antigenic phenotypes in human A(H3N2) influenza viruses.

I present two types of analysis. The first consists of looking for combinations of influenza viruses in the datasets that represent natural versions of experiments we would conduct in the lab. Within this first category I use two slightly different approaches: identifying strains that differ by a substitution, and identifying individually informative strains. The second category uses a LMM to statistically test if molecular changes are associated with antigenic change.

4.2.1 *Strains that differ by single substitutions*

The first type of natural experiment consists of identifying pairs of strains, A and B, that differ by a single substitution, S = XNY. In other words, A and B are genetically identical, except at position N where A has amino acid X, and B has amino acid Y. The antigenic distance between A and B can then be attributed to the substitution S.

A nuance is that there may be multiple strains that are genetically identical to A, so A (and similarly B) can be multiple strains. In fact, when A and/or B are multiple strains, more confidence can be placed in the inferred antigenic impact of S because it is based on the positions of more strains. When multiple strains for A or B are found, their mean position (centroid) is computed and visualised.

Finally, for each substitution, there may be multiple groups of A and B viruses, where different groups have different sequences at non-N positions. Different groups therefore test the impact of the substitution in different genetic contexts. In all analyses different groups are shown on separate antigenic maps. Different groups test the effect of the same substitution but in different genetic backgrounds.

4.2.2 *Individually informative strains*

The second type of natural experiment consists of identifying single strains with informative combinations of amino acids at cluster difference substitution positions. At cluster difference positions, most strains posses the full complement of amino acids that are typical for either the parent cluster, or the child cluster. *Informative strains* have a single position with an amino acid that is atypical of this pattern,

and thus test whether that single molecular change has an antigenic impact.

Consider a case with three cluster difference substitutions: $X_1N_1Y_1$, $X_2N_2Y_2$, $X_3N_3Y_3$. Xs refer to the amino acids possessed by strains in the parent cluster, Ys refer to amino acids possessed by strains in the child cluster, and subscripts index the different substitutions. Most strains have either all Xs at these positions ($N_1X_1 + N_2X_2 + N_3X_3$), or all Ys ($N_1Y_1 + N_2Y_2 + N_3Y_3$). However, if a strain has a single Y amino acid at one position, and Xs at the remaining positions (e.g. $N_1Y_1 + N_2X_2 + N_3X_3$), then it tests the impact of the associated cluster difference substitution.

Specifically, a single Y among many Xs, tests a *forwards* cluster difference substitution, because Xs are amino acids in the parent cluster and Ys are amino acids in the child cluster. Conversely, a single X amino acid among multiple Ys tests a *backwards* cluster difference substitution. For n cluster difference substitutions, there are n forwards, and n backwards possible combinations of strains with informative amino acids.

4.2.3 *LMM association testing*

The final approach for identifying substitutions responsible for antigenic cluster transitions frames the problem as a statistical test in a LMM framework. Section 2.3 contains detailed explanations of this model, data representation, association testing, and multiple test correction.

Briefly, two-dimensional antigenic coordinates are jointly analysed as the response variable in a mvLMM. The amino acid sequence alignment is converted to a binary matrix encoding AAPs at positions 109–301 in HA1 (Subsection 2.3.1). For each cluster transition, all AAPs implicated in cluster difference substitutions are tested. For each test, the test AAP is modelled as a fixed effect and the covariance matrix of the remaining AAPs are treated as random effects. The model jointly estimates the fixed effects in both antigenic dimensions, β_0 and β_1, from which a joint fixed effect is computed:

$$\beta_{joint} = \sqrt{\beta_0^2 + \beta_1^2}$$

A p-value was computed based on likelihoods of models fitted with and without the fixed effect of the test AAP (Subsection 2.3.3). A corrected p-value which takes into account non-independence of AAPs was also computed (Subsection 2.3.4).

4.3 RESULTS

In the following sections I identify cluster difference substitutions for the FU02–CA04, CA04–WI05, PE09–SW13 and PE09–HK14 cluster

transitions (Figure 2.4). Then, for each cluster difference substitution I identify groups of pairs of strains that test the impact of the substitution, identify individual strains with informative combinations of substitutions, and conduct association tests in the LMM framework. Data for the WI05–PE09 cluster transition is not available, so it is treated differently.

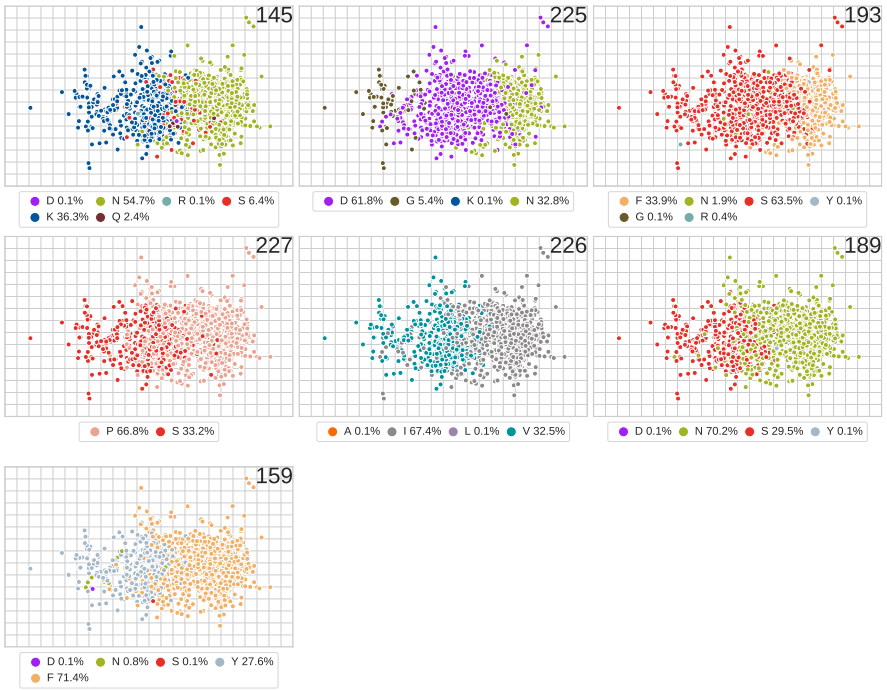


Figure 4.1: Cluster difference substitutions in the Russell et al. (2008b) antigenic map. Strains are coloured according to the amino acid they posses at a given HA position. FU02–CA04 cluster difference substitutions: N145K, S189N, Y159F, V226I and S227P. CA04–WI04 cluster difference substitutions: D225N and S193F.

4.3.1 *FU02–CA04*

Cluster difference substitutions are: N145K, S189N, Y159F, V226I and S227P (Figure 4.1).

STRAINS THAT DIFFER BY SINGLE SUBSTITUTIONS

K145N Three groups were identified (Figure 4.2). (Group 1) The single 145K strain is FU02-like, and the single 145N strain is CA04-like. (Group 2) Seven out of eight 145K strains are FU02-like and the majority of the 84 145N strains are CA04-like. (Group 3) All strains are FU02-like.

Y159F Two groups were identified (Figure E.2). (Group 1) The single 159Y strain is FU02-like. Seven out of eight 159F strains are also

FU02-like, and their centroid is FU02-like. (Group 2) The 44 159Y and five 159F strains are FU02-like.

S189N No strains identified.

V226I 11 groups were identified (Figure E.3). Groups are either all FU02-like (e.g. group 2) or SY97-like (e.g. group 10). No members of any groups are CA04-like.

S227P Nine groups were identified (Figure E.4). (Groups 1, 3, 5) 227P and 227S strains have similar, CA04-like, antigenic locations. (Groups 4, 6) 227P and 227S strains have similar, FU02-like, antigenic locations. (Groups 7–9) 227S strains are FU02-like and 227P strains are CA04-like. (Group 2) Strains differ but are both CA04-like.

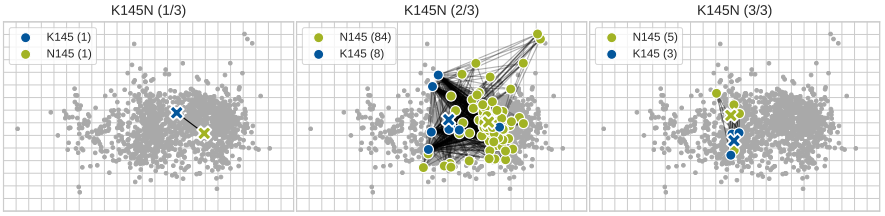


Figure 4.2: Strains that differ by K145N. The three different groups identified for this substitution are shown in separate panels.

INDIVIDUALLY INFORMATIVE STRAINS Of strains that test individual forwards substitutions, none show convincing evidence for causing CA04-like phenotypes (Figure E.6). The two 145N-testing strains remain FU02-like. No strains were identified that test 159F alone. One out of six strains testing 189N are CA04-like, four remain FU02-like and one is intermediate. All 226I- and 227P-testing strains remain FU02-like. Of strains that test individual backwards substitutions, 12 out of 22 identified cause a FU02-like phenotype and six are in-between FU02 and CA04 (Figure E.7). No strains were identified that test either 159Y or 189S. Four out of 15 strains that test 226V are FU02-like and four out of 63 strains testing 227S are FU02-like.

AAP	p-value	Corrected p-value	β_0	β_1	β_{joint}
145N	2.59e-07	9.6e-07	1.94	0.486	2
145K	3.43e-05	0.000127	-1.39	0.568	1.51
159F	0.0137	0.0508	1.35	-0.00135	1.35
226V	0.0221	0.0821	-1.15	-0.568	1.28
159Y	0.101	0.374	0.286	0.792	0.842
189S	0.294	1	-0.819	-0.243	0.854
226I	0.465	1	0.517	0.277	0.587
227S	0.808	1	0.152	-0.189	0.242
227P	0.808	1	-0.152	0.189	0.242
189N	0.867	1	0.226	0.136	0.264

Table 4.3: FU02–CA04 cluster difference substitution LMM association test results. See text for an explanation of each column.

SUMMARY In the pairs-of-strains analyses, K145N group 2 provides the strongest evidence for a substitution that could be responsible for this cluster transition (Figure 4.2). The majority of the eight 145K strains are FU02-like, and the majority of the 84 145N strains are CA04-like. K145N group 1 is also consistent with this. The only other substitution in the pairs-of-strains analyses that is sometimes associated with this cluster transition are S227P groups 7–9. However, these groups collectively contain only four 227S strains and three 227P strains. In the individually-informative analysis, the only forwards-testing strain to become CA04-like are is out of the six 189N strains (another of the 189N strains is in-between FU02 and CA04). Of backwards-testing strains, most 145K strains have a FU02 phenotype, and most 226V and 227S strains are not FU02-like. In the LMM association test, both K145N-testing AAPs have the lowest p-values and the highest joint effect sizes. I conclude that K145N is most likely responsible for this cluster transition.

4.3.2 *CA04–WI05*

Cluster difference substitutions are: D225N and S193F (Figure 4.1).

STRAINS THAT DIFFER BY SINGLE SUBSTITUTIONS

D225N Seven groups were found (Figure 4.3). (Groups 1–7) All groups show low antigenic distance between 225D and 225N strains.

S193F One group was found (Figure 4.3). (Group 1) The majority of the 84 193S strains are CA04-like, although five 193S strains are WI05-like. All eight 193F strains are CA04-like.

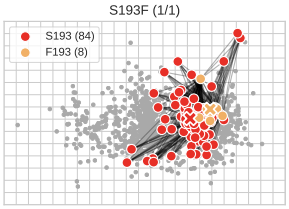


Figure 4.3: Strains that differ by S193F.

INDIVIDUALLY INFORMATIVE STRAINS There are only two cluster difference substitutions for this cluster transition, so strains identified that test forwards substitutions (Figure E.11) are equivalent to those that test backwards substitutions (Figure E.12). Unfortunately, no strains were identified that test the forwards substitution 225N (or backwards substitution 193S). 16 strains were identified showing that 193F produces CA04-like viruses, and simultaneously that 225D does not produce FU02-like viruses.

AAP	p-value	Corrected p-value	β_0	β_1	β_{joint}
193F	0.00321	0.00626	1.06	-0.992	1.45
193S	0.00364	0.0071	-1.02	0.307	1.07
225N	0.314	0.613	0.598	-0.39	0.714
225D	0.874	1	0.107	-0.209	0.234

Table 4.4: CA04–WI05 cluster difference substitution LMM association test results.

SUMMARY D225N does not cause a change from a CA04-like phenotype to WI05-like phenotype in any of the pairs-of-strains analyses. The single group of strains that test S193F do cause this phenotype change (Figure 4.3). Individual strains that possess 193F and 225N are WI05-like. The S193F AAPs have lower p-values and higher joint effect sizes. I conclude that S193F is most likely responsible for this cluster transition.

4.3.3 *WI05–PE09*

The molecular basis of the WI05–PE09 cluster transition cannot be inferred using the same approaches as other cluster transitions in this analysis because no antigenic maps have been published that contain antigens from both clusters. The WI05–PE09 transition straddles the two antigenic maps used in this analysis; the last cluster in the Russell et al. (2008b) map is WI05 and the first cluster in the Li et al. (2016) map is PE09. Conducting an HI assay that includes viruses from both clusters is required to make a map containing both clusters. This is

difficult because no types of red blood cells are known which bind WI05 and PE09 viruses, which is required for running an HI assay containing viruses from both clusters.

Nonetheless, there is circumstantial data which can be used to investigate the molecular determinants of this cluster transition. PE09-like viruses contained two phylogenetically distinct lineages represented by A/Victoria/208/2009 (Victoria clade) and A/Perth/16/2009 (Perth clade) (Klimov et al. 2012). Substitutions were mapped to the separate branches leading to these clades (Klimov et al. 2012). Victoria clade viruses are characterised by the substitutions K158N, N189K and T212A. Perth clade viruses are characterised by E62K, N144K, K158N and N189K. Both sets of substitutions independently generate PE09-like phenotypes, therefore a parsimonious explanation is that the substitutions that are shared between the sets, K158N and N189K, are most likely responsible for this cluster transition.

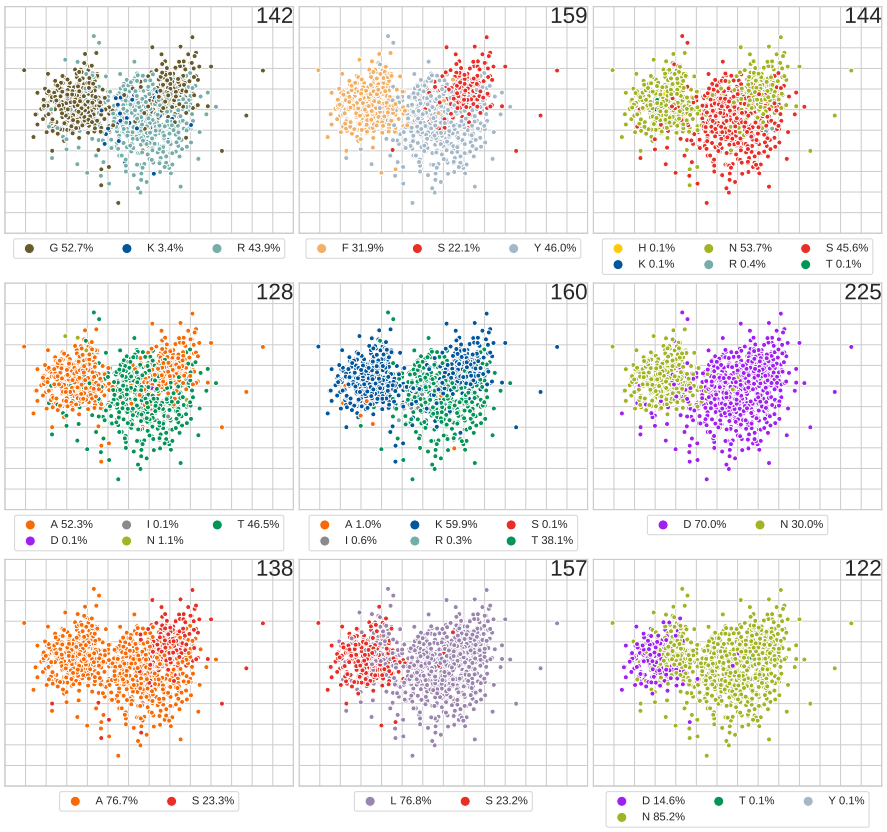


Figure 4.4: Cluster difference substitutions in the Li et al. (2016) antigenic map. Strains are coloured according to their amino acid at a given HA position. PE09–SW13 cluster difference substitutions: A138S, F159S and N225D. PE09–HK14 cluster difference substitutions: A128T, G142R, N144S, F159Y, K160T and N225D.

4.3.4 *PE09–SW13*

Cluster difference substitutions are: A138S, F159S and N225D (Figure 4.4). S157L and D122N partially segregate with PE09-like and SW13-like viruses and are included in this analysis. These two substitutions were included in the pairs-of-strains analyses, but excluded from the individually-informative analysis as their inclusion would reduce the probability of identifying informative strains for the other, clear, cluster difference substitutions.

STRAINS THAT DIFFER BY SINGLE SUBSTITUTIONS

A138S Two groups were identified (Figure E.13). (Group 1) Two out of the three 138A strains remain PE09-like, and one is SW13-like. (Group 2) The only 138S strain is HK14-like.

D122N Two groups were identified (Figure E.14). (Group 1) All strains are PE09-like. (Group 2) All strains are SW13-like.

S157L Three groups were identified (Figure E.15). (Group 1) All strains are PE09-like. (Group 2) Strains are antigenically different, but both PE09-like. (Group 3) Strains are antigenically similar, but also both PE09-like.

F159S One group of strains was identified (Figure E.16). (Group 1) The single 159F strain is PE09-like, whilst the 164 159S strains are SW13-like.

N225D Five groups were identified (Figure E.17). (Groups 1, 2) All strains are PE09-like. (Group 3) All 225N strains are PE09-like, two out of the three 225D strains are also PE09-like and a single 225D strain is SW13-like. (Group 4) The 225N strain is PE09-like and the 225D strain is intermediate between the three clusters. (Group 5) The 225N strain is in-between PE09 and SW13 and the 225D strain is in-between SW13 and HK14 strain.

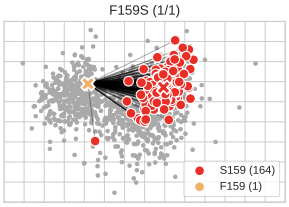


Figure 4.5: Strains that differ by F159S.

INDIVIDUALLY INFORMATIVE STRAINS Strains were only identified that test one forwards substitution (Figure E.19). The majority of the 16 225D-testing strains are PE09-like, whilst one is SW13-like. Similarly, strains were only identified that test one of the backwards

substitutions (Figure E.20). Both of the two 159F-testing strains are PE09-like.

AAP	p-value	Corrected p-value	β_0	β_1	β_{joint}
159F	1.91e-15	7.48e-15	-3.16	-0.168	3.16
159S	2.52e-13	9.85e-13	1.53	1.5	2.15
225D	0.00103	0.00403	0.136	-1.33	1.34
225N	0.00408	0.016	-0.414	0.96	1.05
138A	0.0353	0.138	-1.34	-0.382	1.39
122N	0.0731	0.286	0.948	0.362	1.01
122D	0.215	0.84	-0.691	-0.376	0.786
138S	0.459	1	-0.528	-0.468	0.706
157S	0.836	1	-0.235	-0.187	0.3
157L	0.836	1	0.235	0.187	0.3

Table 4.5: PE09–SW13 cluster difference substitution LMM association test results.

SUMMARY There are only three groups in the pairs-of-strains analyses that show substitutions associated with the PE09–SW13 transition: F159S (group 1), N225D (group 3) and A138S (group 1). For N225D and A138S, for each of their single SW13-like strains there are two genetically identical strains that remain PE09-like, suggesting that the strains with SW13-like phenotypes are erroneous. However, there is only a single 159F substitution to support 159F strains being PE09-like. Only 225D had strains identified in the individually-informative forwards-testing analysis; 15 out of 16 strains remain PE09-like and a single strain becomes SW13-like. In the backwards-testing analysis, both 159F strains do revert to being PE09-like. In the LMM association test the F159S AAPs have the lowest p-values and the largest joint effect sizes. I conclude that F159S is most likely responsible for this cluster transition.

4.3.5 *PE09–HK14*

Cluster difference substitutions are: A128T, G142R, N144S, F159Y, K160T are N225D (Figure 4.4). D122N and S157L co-segregate partially with either PE09-like or HK14-like viruses but are nonetheless included in this analysis. As for the PE09–SW13 analyses, these two substitutions were included in the pairs-of-strains analyses, but excluded from the individually-informative analysis as their inclusion would reduce the probability of identifying informative strains for the other, clear, cluster difference substitutions.

STRAINS THAT DIFFER BY SINGLE SUBSTITUTIONS

D122N Two groups were identified (Figure E.21). (Groups 1, 2) No strains are HK14-like.

A128T One group of strains was identified (Figure E.22). (Group 1) The centroids of the 128T strains are antigenically similar to the 128A strain.

G142R One group of strains was identified (Figure E.23). (Group 1) The two 142G strains are HK14-like and the majority of the 142R strains are either SW13-like of HK14-like.

N144S No strains identified.

S157L See Subsection 4.3.4 and Figure E.24. There is no evidence that this substitution causes this cluster transition.

F159Y Three groups were identified (Figure E.25). (Group 1) All 159F strains are PE09-like, and the 159Y strain is in-between PE09 and HK14. (Group 2) All 159F strains are PE09-like, and the 159Y strains are at the top of HK14. (Group 3) The single 159F strain is PE09-like, two out of three 159Y strains are HK14-like and the other is SW13-like.

K160T Six groups were identified (Figure E.26). (Groups 1, 2, 4, 5) The centroids of the 160K and 160T strains are antigenically close and there is no clear change from PE09-like to HK14-like. (Group 3) The 160K strain is HK14-like, and the 160T strain is PE09-like. This is the opposite direction to the change in the majority of strains with these amino acids at 160 (Figure 4.4). (Group 6) The 160T strain is on the edge of the HK14 cluster and the 160K strain is on the outskirts of the map in-between the PE09 and HK14 clusters.

N225D See Subsection 4.3.4 for a complete description and Figure E.27. (Group 4) 225N is PE09-like and 225D is HK14-like.

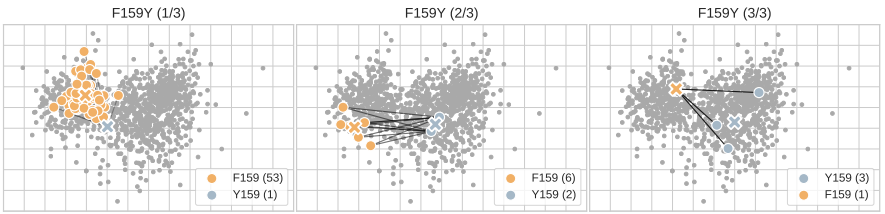


Figure 4.6: Strains that differ by F159Y.

INDIVIDUALLY INFORMATIVE STRAINS Only 159Y, 144S and 225D had strains which tested their individual forward effects (Figure E.29). None of the strains were HK14-like. 142G, 159F, 128A and 160K had strains which tested their individual backwards effects (Figure E.30). The single 159F strain reverts to being PE09-like, as does one of the 32 160K strains.

AAP	p-value	Corrected p-value	β_0	β_1	β_{joint}
159F	1.91e-15	9.72e-15	-3.16	-0.168	3.16
160T	1.08e-11	5.51e-11	0.777	0.648	1.01
159Y	0.000617	0.00314	0.628	-1.1	1.27
225D	0.00103	0.00523	0.136	-1.33	1.34
225N	0.00408	0.0207	-0.414	0.96	1.05
160K	0.0194	0.0986	-0.33	0.428	0.54
122N	0.0731	0.372	0.948	0.362	1.01
142G	0.18	0.913	-0.401	-0.782	0.878
122D	0.215	1	-0.691	-0.376	0.786
144S	0.233	1	-0.696	-0.326	0.769
142R	0.317	1	-0.37	0.387	0.536
128T	0.482	1	-0.491	-0.222	0.539
128A	0.68	1	0.0594	0.35	0.355
157S	0.836	1	-0.235	-0.187	0.3
157L	0.836	1	0.235	0.187	0.3
144N	0.96	1	-0.0453	0.0894	0.1

Table 4.6: PE09–HK14 cluster difference substitution LMM association test results.

SUMMARY In the pairs-of-strains analyses two out of three F159Y groups and two out of five N225D groups show a change from PE09-like viruses to HK14-like viruses. For individually informative strains, the single 159Y forwards-testing strain is in-between PE09 and HK14. The single 144S forwards-testing strain, and majority of the 225D forwards-testing strains, remain PE09-like. For backwards-testing strains only the single 159F strain, and one out of the 32 160K strains become PE09-like. The F159Y AAPs have the first and third lowest p-values in the LMM association test. 160T has the second lowest p-value, however the other AAP in this substitution, 160K, has the sixth lowest p-value. I conclude that F159Y is most likely responsible for this cluster transition.

4.4 DISCUSSION

In this chapter I reviewed substitutions that are known to cause major antigenic change in influenza viruses from a variety of subtypes and hosts, and identified substitutions responsible for antigenic cluster transitions in human A(H3N2) influenza viruses that have circulated since the early 2000s. The methods I used to identify these substitutions differ from those in the work I reviewed (Lewis et al. 2011; Koel et al. 2013; Koel et al. 2014; Lewis et al. 2014; Abente et al. 2016), After identifying candidate substitutions, these studies used reverse genetics to systematically test whether single, or combinations of multiple, substitutions were responsible for an antigenic change. I used GISRS data and looked for natural versions of the same experiments, as well as conducting statistical association tests.

Both approaches have benefits. The reverse genetics approach gives precise control of the substitutions and genetic background of the mutant viruses, which reduces the risk of having no data for a particular substitution. It does not remove this risk entirely; occasionally influenza virus mutants cannot be rescued in the laboratory. The reverse genetics approach uses small panels of well characterised antisera and reference antigens to measure antigenic phenotypes. The associated antigenic maps can be more accurate and contain better coordinated antigens than in GISRS maps. Looking for natural experiments also has benefits: it leverages the substantial amount of data that is generated by the GISRS and substitutions are tested in a variety of genetic backgrounds. These approaches are highly complementary. Natural experiments could be a more refined way to identify substitutions to test in the laboratory. The time and resources saved in screening additional candidate substitutions could be spent testing substitutions in a greater variety of genetic backgrounds.

The potential importance of testing substitutions in different genetic backgrounds is illustrated by the K145N pairs-of-strains analyses (Figure 4.2). Group two robustly shows that strains with 145K are antigenically different to strains with 145N. This difference is supported by group one, albeit by only two strains. However, in group three, the 145N strains remain FU02-like, near to the 145K strains. One hypothesis explaining that an antigenic difference is observed in groups one and two but not in group three is that the different genetic background alters the antigenic impact of this substitution. This case merits additional laboratory verification, but it may comprise an interesting test case to investigate structural mechanisms of antigenic context dependence.

One drawback of using GISRS derived data is the higher amount of noise in the antigenic coordinates (discussed in Chapter 1 and Chapter 2). The A138S pairs-of-strains analyses highlights this (Figure E.13). Two out of three strains with 138A remain PE09-like, and the other

strain becomes SW13-like. I interpreted the position of the SW13-like strain to be erroneous because it is outnumbered by genetically identical PE09-like strains two:one. Clearly, inferences based on more strains are more robust.

Repeating these analyses on sanitised maps would remove these judgements (Appendix A). In sanitised maps genetically identical strains are also antigenically identical. This would ease the interpretation of the natural experiments where genetically identical antigens can be antigenically very different. Currently, map sanitising has not been fully tested, and it is unknown if it is a valid post-processing step for antigenic maps.

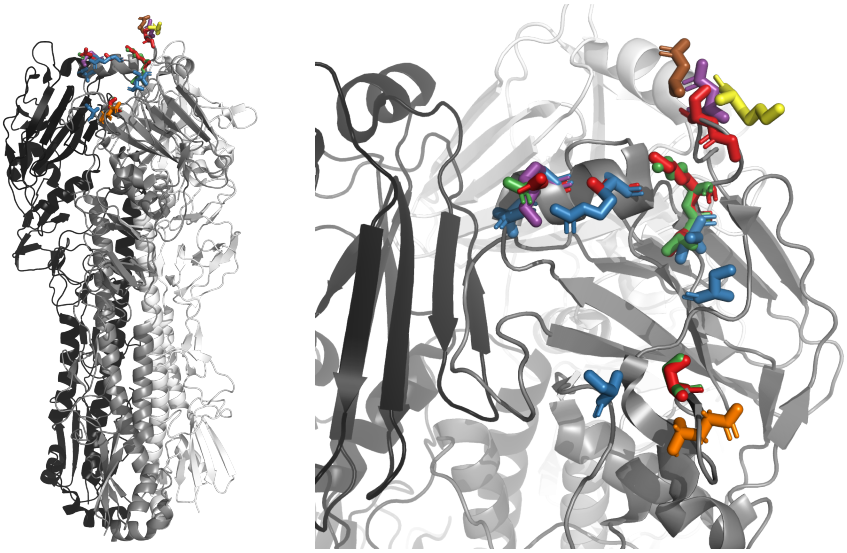
Despite some limitations of the natural experiments approach, it allowed robust inferences of the molecular basis of antigenic change. These inferences are made more robust in combination with the LMM association tests which corroborate the substitutions identified.

Substitutions identified in this chapter continue trends in properties of other human A(H3N2) cluster transition substitutions (Table 4.7). In particular, all the substitutions occur in the same seven positions on the periphery of the HA RBS. Furthermore, cluster transition substitutions in other influenza viruses, reviewed at the start of this chapter, also occur on the periphery of the HA RBS (Figure 4.7). The spatial clustering of these substitutions around the rim of the HA RBS suggests a shared mechanism of antigenic escape across diverse influenza viruses: antibodies target the RBS region and so this region must change to provide escape from host immune surveillance. However, the HA must simultaneously retain receptor binding functionality which is critical for virus replication. Such a process implicates constraints on the nature of amino acid substitutions that cause antigenic change in influenza viruses.

In the next chapter I quantitatively investigate features of cluster transition substitutions, and test to what degree the features repeat. I then apply the findings to ranking sets of candidate substitutions by their similarity to previous cluster transition substitutions to aid the prediction of antigenic evolution.

Substitution	Cluster transition
K145N	FU02–CA04
S193F	CA04–WI05
K158N, N189K	WI05–PE09
F159S	PE09–SW13
F159Y	PE09–HK14

Table 4.7: Cluster transition substitutions in human seasonal A(H3N2) influenza viruses isolated between 2004 and 2016.



(a) Entire trimer. (b) RBS region.

Figure 4.7: HA positions that harbour cluster transition substitutions in different influenza virus subtypes and hosts. Individual monomers are shown in white, grey and black. On the grey monomer, positions where cluster transition substitutions occurred are rendered with sticks and shown in different colours for different influenza viruses: red, human A(H_3N_2); blue, avian A(H_5N_1); green, swine A(H_3N_2); purple, equine A(H_3N_8); orange, human A(H_1N_1); yellow, human B(Victoria); brown, human B(Yamagata). This figure was rendered using PyMOL (Schrödinger LLC 2015) on the A/Hong Kong/1/1968 (H_3N_2) crystal structure, PDB ID: 4FNK.

5 CLUSTER TRANSITION SUBSTITUTION RANKINGS

CHAPTER OUTLINE This chapter describes the development of a framework to rank candidate substitutions by their similarity to cluster transition substitutions. Koel et al. (2013) noted the pronounced patterns in biophysical features of cluster transition substitutions. In this chapter I quantify these patterns, and demonstrate that the processes generating them are highly unlikely to be random. More fundamentally, work in this chapter investigates the repeatability of these patterns, and tests whether they can be used to predict cluster transition substitutions.

5.1 INTRODUCTION

This work is motivated from two angles. First is the basic science of testing the predictability of evolution. Seasonal influenza viruses are an excellent model system to investigate this question. Influenza viruses have attracted substantial research effort, resources and surveillance which has driven understanding in their virology, immunology, epidemiology and structural biology for almost a century. Furthermore, they evolve on a timescale which allows predictions to be prospectively tested.

Second are the implications for vaccination strategies. Currently, most countries use the WHO choice for influenza vaccine strains. Due to manufacturing, logistic and regulatory constraints, vaccine strain choices must be made approximately eight months before the start of the influenza season (Subsection 1.5.1). Cluster transitions may occur during this period, and if they do, the vaccine would be antigenically mismatched to viruses in the upcoming season. Mismatch events could be ameliorated by predicting future antigenic variants and applying a novel vaccination strategy based on the concept of *immunity management*.

5.1.1 *Immunity management*

Unlike the current WHO vaccination strategy which selects vaccine strains to antigenically match circulating strains, immunity management prioritises generating population immunity ahead of the antigenic evolution of the virus. If an antigenic cluster has already circulated, then instead of repeatedly using the same vaccine, this opportunity could be used to vaccinate against future variants. If we predict a particular antigenic variant, and that variant does go on to dominate

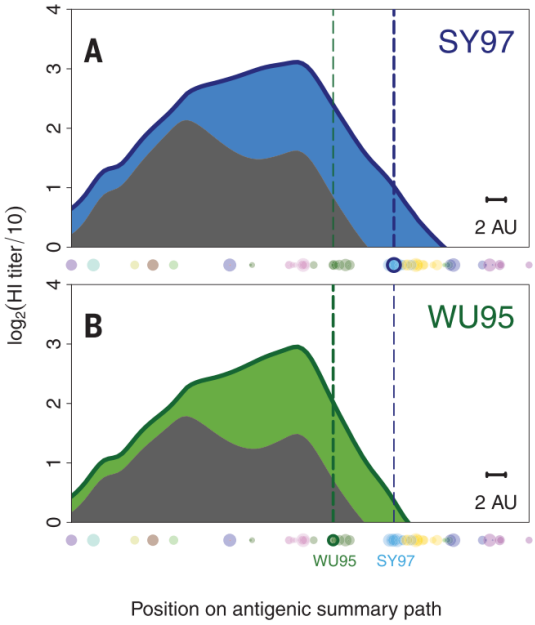


Figure 5.1: Comparing the immune response to two vaccines. (A) SY97 vaccine. (B) WU95 vaccine. Plots show the height of the population antibody landscape (y-axis), along a one dimensional summary path through antigenic space (x-axis). Grey areas show pre-vaccination antibody landscapes, coloured areas show the post-vaccination landscape. The response to each vaccine is shown by the height of the coloured area above the grey area. Green and blue dotted vertical lines show positions of the WU95 and SY97 vaccines. The SY97 vaccine is more antigenically advanced. Both vaccines generate a similar back-boost but SY97 generates a higher response against antigenically advanced strains. This is Figure 3 from Fonville et al. (2014); it is based on empirical data from vaccine trials.

subsequent circulation, then we have successfully avoided a mismatch event. However, what are the consequences if either no drift variants circulate, or, if a variant different from that predicted circulates?

In the scenario where no drift variant circulates, evidence from *antibody landscape* suggests the advanced strain would not be inferior to a matched strain (Fonville et al. 2014). In an antibody landscape, an individuals antisera is titrated against different viruses across antigenic space. Then, a surface is fitted through the titrations, in an extra dimension above the antigenic map. Antibody landscapes therefore provide quantitative and visual insight into an individual's immune response against viruses across antigenic space. Antibody landscapes revealed that influenza infection or vaccination induces a broad response against all strains an individual has been exposed to in their lifetime (Fonville et al. 2014). This phenomenon is referred to as the *back-boost*. Thanks to the back-boost, an antigenically advanced

strain generates the same protective responses against previously encountered strains as an antigenically matched strain.

Empirical evidence supports this reasoning. In 1997 an antigenically novel variant, A/Sydney/5/1997 (SY97), emerged, and the vaccine strain was updated accordingly. However, there was very low circulation of SY97-like viruses in Australia during the subsequent season. Australians had effectively received an antigenically advanced vaccine. Individuals that received the SY97 vaccine had the same back-boost as individuals vaccinated in the previous year with the previous vaccine. Importantly, the SY97 vaccine also generated higher titres against the novel SY97-like viruses (Figure 5.1).

In the other scenario, major antigenic evolution occurs, but an incorrect variant is predicted. Here again, antigenically advanced strains would not be inferior to matched strains, as both vaccines would be mismatched. Clearly, we aim to reduce the occurrence of this case, and doing so requires improving our ability to predict antigenic evolution.

5.1.2 *Combinations of antigenic escape variants*

Our ability to approach predicting the antigenic evolution of human A(H3N2) viruses is based on a detailed understanding of the molecular basis of their antigenic change, and a relatively low number of potential variants to screen. As our knowledge has improved, this number has shrunk.

Work in the 1980s suggested that major antigenic change required at least four amino acid substitutions located within 131 HA amino acid positions (Wiley et al. 1981; Wilson and Cox 1990). These 131 positions were divided into five HA surface patches, termed antigenic sites A-E. The work further implied that major antigenic change required the four substitutions to not all occur in a single antigenic site.

Given these 'rules', it is possible to compute how many combinations of substitutions might cause major antigenic change. Antigenic sites A–E contain 19, 22, 27, 41 and 22 HA positions, respectively. The number of combinations of four substitutions distributed between these antigenic sites, excluding combinations where all four occur in a single antigenic site is:

$$\binom{131}{4} - \binom{19}{4} - \binom{22}{4} - \binom{27}{4} - \binom{41}{4} - \binom{22}{4} = 11,579,314$$

A substitution is a change from one amino acid to one of 19 other possibilities, giving:

$$19^4 = 130,321$$

combinations of four substitutions. Each combination of four substitutions could be distributed in any way among the combinations of HA positions, giving:

$$11,579,314 \times 130,321 = 1.51 \times 10^{12}$$

total combinations. This number is large enough to prevent any efforts attempting to predict which combination might cause future antigenic change. Furthermore, these 'rules' were recognised as the minimum required; some cases required more than four substitutions, which would increase the number of combinations.

In 2013, the substitutions that caused antigenic cluster transitions in human seasonal A(H3N2) viruses that circulated between 1968 and 2003 were identified (Koel et al. 2013). Remarkably, all substitutions identified occurred at only seven key positions on the periphery of the HA RBS, and seven out of ten cluster transitions were caused by single substitutions. If all antigenic cluster transitions are caused by single amino acid substitutions at a Koel-seven position, then only:

The HA positions responsible are 145, 155, 156, 158, 159, 189 and 193 (Koel et al. 2013). They are sometimes referred to as Koel-seven positions.

$$19 \times 7 = 133$$

variants would have to be discriminated between to identify which will cause a subsequent cluster transition. There are:

$$\binom{7}{2} \times 19^2 = 7,581$$

combinations of double substitutions distributed between seven positions.

Combinations of single (133) and double (7,581) variants to discriminate between are eight and nine orders of magnitude smaller than the 1980s estimate of 10^{12}, respectively. These data suggest that if historical patterns repeat themselves, then in seven out of ten cases, cluster transition substitutions may be predictable by discriminating between only 133 influenza virus variants. The relatively small number of variants to discriminate between, coupled with their utility in immunity management vaccination strategies is a strong motivator for the ranking work in this chapter.

5.1.3 *Virus fitness*

Using current technology, a single researcher can generate an influenza virus library containing all single amino acid variants at seven amino acid positions in approximately three months. Viruses in the library can then be subjected to assays which measure intrinsic and extrinsic fitness (Subsection 1.4.1). Fitness measurements are a crucial component of a framework to predict antigenic variants and are discussed in Chapter 6.

5.2 PATTERNS IN CLUSTER TRANSITION SUBSTITUTIONS

Cluster transitions substitutions in human A(H3N2) influenza viruses identified by Koel et al. (2013) have features that appear to be non-random: some amino acids occur frequently, the charge of the residue usually changes, hydropathy of residues often remains similar, change in residue side chain volume is variable, they are usually achievable by mutating a single nucleotide and particular positions around the HA RBS are repeatedly used. I defined these features and quantified how likely it is that these patterns were generated by a random process. Amino acid substitution features were defined as follows.

5.2.1 *Categorical features*

- The amino acid lost in the substitution. Abbreviated *aa0*.
- The amino acid gained in the substitution. Abbreviated *aa1*.
- Charge change. Any substitution that causes any change in charge was encoded as 1, all others are encoded with a 0. The charge of amino acids at physiological pH (7.4) was used. Negative amino acids are glutamic acid and aspartic acid, positive amino acids are lysine, arginine and histidine, all others are neutral. Abbreviated *dC*.
- Minimum mutation distance. Each codon has a minimum number of mutations to mutate into another codon—the Hamming distance between the two codons. The minimum mutation distance of an amino acid substitution is the lowest Hamming distance between the two sets of codons that encode the two amino acids. Abbreviated *mmd*.

5.2.2 *Continuous features*

- The HA position that the substitution occurs at. This is represented by 3D Cartesian coordinates of the α-carbon of the HA residue. This representation has two advantages compared to treating each HA position categorically. Firstly, positions on a protein are not spatially-independent: observing a cluster transition substitution at one position should raise the probability of observing a future cluster transition substitution at nearby positions. This is consistent with the observation that cluster transition substitutions occur near to each another in 3D protein structure (Koel et al. 2013). An ordinal treatment where HA positions are represented by their position in the primary sequence is inferior because positions far away in protein primary structure can be close in tertiary and quaternary structure.

Secondly, later in this study, I investigated using training data from other influenza viruses (Subsection 5.3.7). Occasionally it is impossible to identify unambiguously homologous positions in influenza HA structures from different types, subtypes and hosts due to backbone shifts, insertions and deletions. Representing HA position continuously circumvents this issue, which would be required for a categorical representation. Abbreviated *xyz*.

- Hydropathy and volume change. Hydrophobicity is a measure of the tendency of a molecule to be repelled by water and is measured on hydropathy scales (Kyte and Doolittle 1982). The hydropathy change of substitution, s, from amino acid, $aa0$, to amino acid, $aa1$, is:

$$\Delta H_s = H_{aa1} - H_{aa0}$$

where H is hydropathy corrected for effects of self-solvation between side chain and backbone atoms of a particular amino acid residue (Roseman 1988). Similarly, the volume change of a substitution is:

$$\Delta V_s = V_{aa1} - V_{aa0}$$

where V is the normalised Van der Waals volume of the amino acid side chain (Fauchère et al. 1988). The joint distribution of hydropathy and volume changes was used as a feature. Abbreviated *dHdV*.

I generated network diagrams to visualise sets of substitutions and their biophysical features (Figure 5.2). These diagrams consist of amino acids as nodes in a hydropathy-volume plot where arrows between nodes visualise substitutions from one amino acid to another. Charge change and minimum mutation distance are also visualised in the plot. Human A(H3N2) influenza virus cluster transition substitutions contain interesting patterns (Figure 5.2D). Lysine is a heavily used amino acid—of the 14 substitutions in only six do not involve lysine in some way. In addition, the majority of substitutions comprise amino acids that have similar values of hydropathy and volume as lysine. 11 out of the 14 substitutions involve a charge change, and two out of the three that do not involve a charge change cause a large change in side chain volume (TY, and SY). TY is the only substitution that does not have a minimum mutation distance of one.

5.2.3 *Empirical p-values of categorical features*

Next I computed empirical p-values to measure the probability of observing the patterns in Figure 5.2D, or more extreme patterns, if amino acids in substitutions were selected at random. I did this firstly

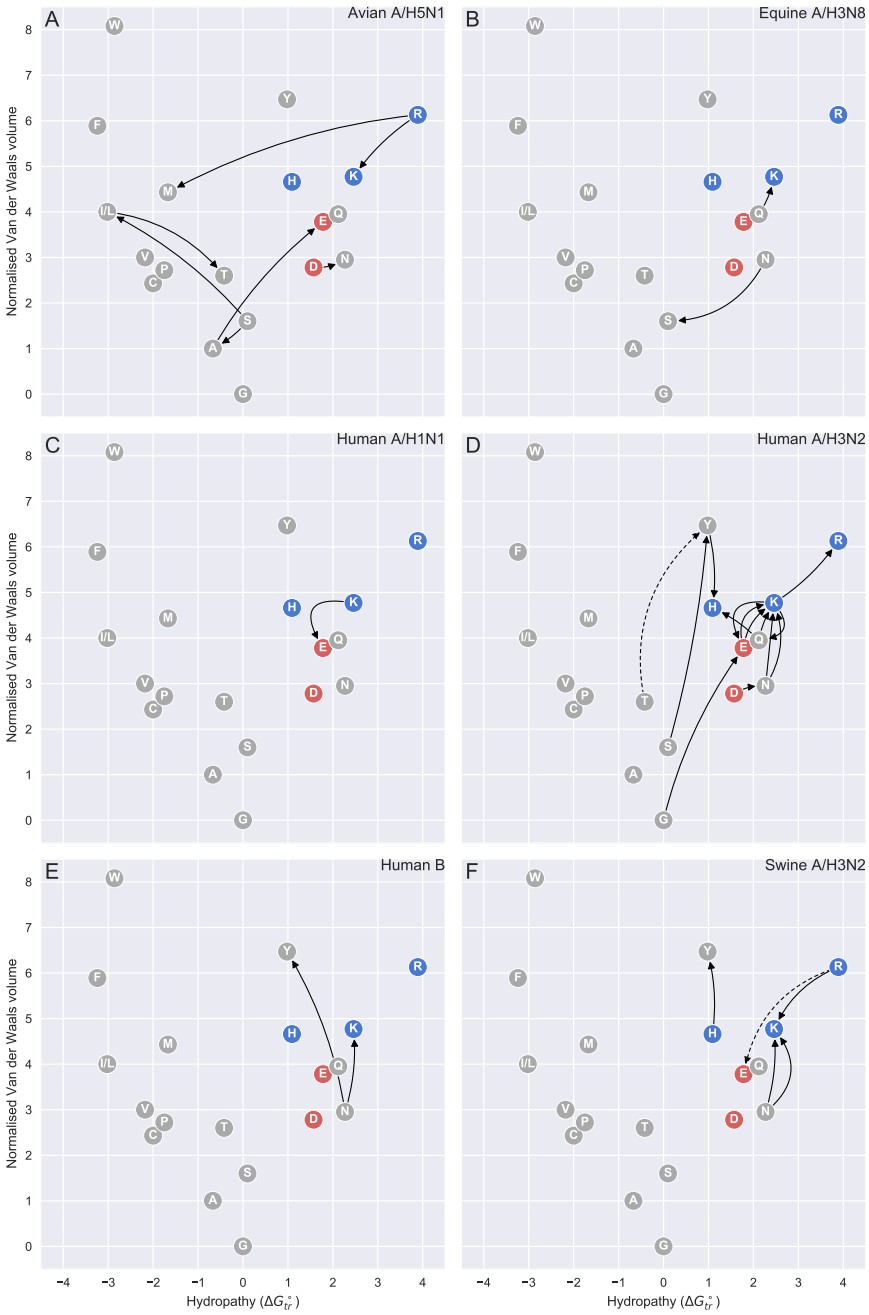


Figure 5.2: Cluster transition substitutions in different influenza viruses. (A), Avian A(H5N1); (B), Equine A(H3N8); (C), Human A(H1N1); (D), Human A(H3N2); (E), Human B/Victoria and B/Yamagata; (F), Swine A(H3N2). Amino acids are points plotted according to their hydropathy and normalised Van der Waals volume. Substitutions are arrows between amino acids. Negatively and positively charged amino acids are red and blue, respectively. Substitutions with a minimum mutation distance of two have dashed arrows, those with a value of one have solid arrows. (No substitutions in this figure have a minimum mutation distance of three.)

to quantify the qualitative descriptions of these patterns, and secondly to act as inclusion criteria for features in the ranking framework (Section 5.3).

A standard way to test whether counts in categorical data deviate from expected frequencies is to use a χ^2 test. This test is invalid when counts in each category are low; an often used heuristic requires a minimum count of five in each category. These data violate this heuristic (Table 5.1), so instead I computed empirical p-values.

For each feature, the probability of the observed data, $P(D_o)$, was computed using the multinomial distribution. The multinomial distribution gives the probability of observing counts, D_o, given probabilities, p, for each category in the data. For example, consider the aa1 data of human A(H3N2) cluster transition substitutions (Table 5.1). There are 20 standard amino acids, so for aa1, a uniform probability distribution is a length 20 vector: $p = (\frac{1}{20}, ..., \frac{1}{20})$. Substituting the aa1 counts and a uniform probability distribution into the multinomial probability mass function gives:

An extension of this analysis could use a null distribution derived from empirical data. For instance, the observed frequencies of amino acids in human A(H3N2) viruses could be used to compute p.

$$P(D_o) = 5.54 \times 10^{-11}$$

This is the probability of observing exactly these aa1 counts given a uniform probability distribution. We are actually interested in knowing what the probability is that a random process would generate this, or even more extreme, data. This is equivalent to computing how likely it is that randomly generated data has a lower probability than $P(D_o)$. To estimate this, $n = 1 \times 10^7$ sets of 14 random substitutions were generated by selecting pairs of amino acids from a uniform random distribution. 14 is used here because there are 14 human seasonal A(H3N2) cluster transition substitutions in the Koel et al. (2013) dataset. The probability of each sampled dataset, $P(D_i)$, was computed using the multinomial distribution and a uniform probability distribution. The proportion of times that $P(D_i)$ was less than $P(D_o)$ is the empirical p-value:

$$p_e = c/n \tag{5.1}$$

where c is the number of sampled datasets with $P(D_i) < P(D_o)$.

5.2.4 *Empirical p-values of continuous features*

Empirical p-values were also computed for continuous features. As for categorical features, I sought to quantify how frequently a dataset more extreme than the observed dataset is generated if amino acids were selected at random. For categorical data *extremeness* was quantified using the multinomial likelihood function and a uniform probability distribution, which cannot be applied to continuous data.

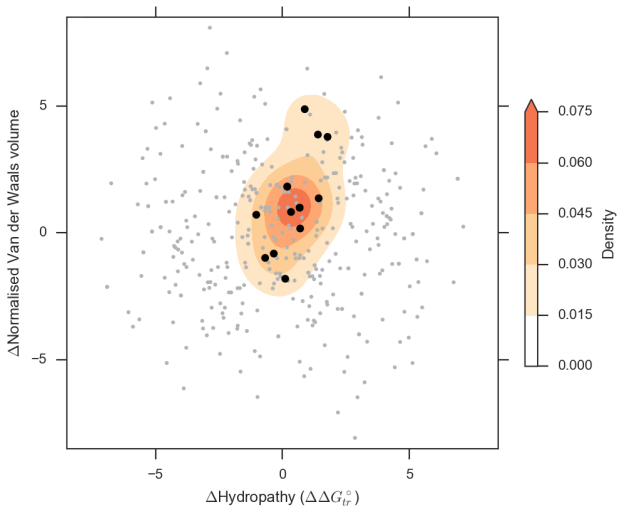


Figure 5.3: KDE of hydropathy and volume changes of amino acid substitutions. Koel et al. (2013) cluster transition substitutions are shown in black, all other possible substitutions are shown in grey. Contours show KDE density of the cluster transition substitution.

Instead, the *extremeness* of a continuous feature from a set of substitutions, s, was quantified as the likelihood of that feature for all 380 possible substitutions given a model trained on s, M_s:

$$\mathcal{L}_{All|M_s} = \prod_{j}^{380} P(D_j|M_s) \tag{5.2}$$

KDE was used as the model. I used the SciPy KDE implementation (Jones et al. 2001), Gaussian kernels, and Scott's rule to estimate bandwidths (Scott 1992).

As for categorical features, $n = 1 \times 10^7$ sets of 14 substitutions were generated by selecting amino acids at random. For each set of 14 random substitutions, r, the likelihood of observing all substitutions given a model trained on r was computed, $\mathcal{L}_{All|M_r}$. The empirical p-value was then computed using Equation 5.1, where c is the number of sampled datasets where $\mathcal{L}_{All|M_r} < \mathcal{L}_{All|M_{D_o}}$.

Empirical p-values corroborate the qualitative description of patterns in amino acid features of cluster transition substitutions outlined above (Table 5.1). Only data from aao is consistent with being generated from a random process.

5.3 A FRAMEWORK TO RANK CLUSTER TRANSITION SUBSTITUTIONS

I will now describe the development of a framework to rank candidate substitutions by their similarity to historic cluster transition substitutions. The aim is to predict which out of a set of candidate

Feature	Data	p_e
aao	K,3; N,2; Q,2; E,2; D,1; S,1; G,1; Y,1; T,1; A,0; C,0; F,0; H,0; I,0; L,0; M,0; P,0; R,0; V,0; W,0	0.196
aa1	K,5; H,2; E,2; Y,2; R,1; Q,1; N,1; A,0; C,0; D,0; F,0; G,0; I,0; L,0; M,0; P,0; S,0; T,0; V,0; W,0	0.00282
dC	0,3; 1,11	0.00702
mmd	1,13; 2,1; 3,0	0.000182
dHdV	(See Figure 5.2 and Figure 5.3.)	0.0000008

Table 5.1: Features and empirical p-values of Koel et al. (2013) cluster transition substitutions.

substitution(s) will cause the next cluster transition. *Candidate substitutions* refer to substitutions that are possible given currently circulating human seasonal A(H3N2) influenza viruses.

In this section I will outline: the formulation of a metric, h_s, that is used to rank substitutions; the computation of probability distributions for substitution features; the generation of candidate substitution sets; the development of the framework using one set of cluster transition substitutions, and testing the framework using a separate, blinded test set; bespoke visualisations (*bumps plots*) for visualising how rankings change as training data accumulate and as the HA evolves; testing whether additional training data improves rankings; and results from relaxing the assumption that future cluster transition substitutions will occur at a Koel-seven position.

5.3.1 *A score to rank substitutions*

This is a *learning to rank* problem with the goal of ranking candidate substitutions by a score related to the likelihood that they will cause a cluster transition. Due to the low amount of data available and without a mechanistic understanding of what governs the molecular antigenic evolution of human A(H3N2) seasonal influenza viruses, I do not aim to accurately estimate this likelihood. Instead, I aim to take sets of substitutions that have caused cluster transitions (*training substitutions*), and make a framework that scores candidate substitutions by their similarity to these training substitutions. This approach was motivated by the pronounced patterns observed in features of human A(H3N2) influenza virus cluster transition substitutions (Section 5.2).

One issue related to the low amount of training data is that there is an absence of data for what cluster transition substitutions are not like. When developing classification systems it is useful to have positive examples (usually represented as 1s) as well as negative examples (usually represented as 0s) in the training data. This partially

motivated the design of the score, h_s, where the score of substitution, s is simply the product of the probabilities of its features:

$$h_s = \prod_i P(x_{is}) \tag{5.3}$$

where x_{is} denotes the value of feature i in substitution s and $P(x_{is})$ derives from probability density functions parametrised from a set of training substitutions (Subsection 5.3.2). h_s for a set of candidate substitutions are normalised such that they sum to one.

This formulation of h_s assumes features are independent, which strictly, is likely to be untrue for these data. This formulation is a simple technique to quantitatively combine evidence from multiple features *in lieu* of more data which could support a more sophisticated model. The same treatment is used by naïve Bayes classifiers, which generally perform well in cases, such as this one, with low amounts of training data (Murphy 2012).

Two events, A and B, are independent if the probability of observing both is equal to the product of the probabilities of observing either alone: $P(A \cap B) = P(A)P(B)$.

I excluded features which may have been generated by random processes. Therefore, I use features with empirical p-values lower than five percent: aa1, dC, mmd (Table 5.1), as well as HA position, for which I did not conduct a statistical test. I note that all the features with empirical p-values below this threshold would also be below a threshold one order of magnitude lower.

5.3.2 *Probability density functions of amino acid substitution features*

Probability distributions for feature values are required to compute h_s (Equation 5.3). Categorical and continuous features were handled separately.

Data for categorical features comprise an array of counts, D, corresponding to the number of times each value is observed in data. Bayesian inference was used to derive probability distributions for categorical features using the multinomial likelihood function. A Dirichlet distribution, parametrised by an array, α, was used for the prior, $\mathrm{Dir}(\alpha)$. Plus one smoothing was used due to the low amount of training data (Murphy 2012), which was implemented by using $\alpha = (1, ..., 1)$. Due to multinomial-Dirichlet conjugacy, the posterior distribution is also a Dirichlet distribution, $\mathrm{Dir}(\alpha + D)$. Marginal posterior distributions represent the probability distributions for each category of the feature. Marginal distributions of a Dirichlet distribution, $\mathrm{Dir}(a)$, are Beta distributions, $\mathrm{Beta}(a_i, \sum a - a_i)$, which were summarised by computing the mean.

Probability distributions of continuous features were modelled using KDE. The scikit-learn implementation with Gaussian kernels was used (Pedregosa et al. 2011). When the number of data points was greater than six, optimal bandwidths were determined via three-fold cross-validation. Otherwise, a suitable bandwidth was selected using Scott's rule (Scott 1992).

5.3.3 *Candidate substitutions*

For a particular antigenic cluster, rankings were generated by computing h_s for all candidate substitutions given a set of training substitutions, and then sorting the candidates by h_s. Candidate substitutions were generated from the consensus HA sequence of viruses in an antigenic cluster. As the HA, evolves the set of candidate substitutions that is being ranked changes. For example if an asparagine (N) is present at a particular position in one cluster, the candidate substitutions for that position will constitute N to the 19 other possible amino acids. In the next cluster, if that position has evolved into cysteine (C), then the candidate substitution for that position will now constitute C to the 19 other possible amino acids. Initially candidate substitutions were generated at Koel-seven positions only. Later, I extended the candidate substitutions to include all surface exposed HA1 amino acid residues less than 35 Å from a Koel-seven residue (Subsection 5.3.10).

5.3.4 *Training data*

Training data for each ranking consisted of only human A(H3N2) cluster transition substitutions from preceding cluster transitions. For example, the three clusters prior to TX77 are HK68, EN72 and VI75 so, only cluster transition substitutions from the HK68–EN72, EN72–VI75 and VI75–TX77 transitions were used to rank the TX77 candidate substitutions. Omitting training data from subsequent cluster transitions was done in order to better generalise the performance of this ranking framework into the future.

I omitted BE89 which is thought to be an evolutionary dead-end, thus has no cluster transition substitutions derived from it, and therefore has no training data to contribute.

The SI87 and BE92 antigenic clusters were also omitted. These are the two clusters for which influenza virus libraries containing all single amino acid variants at Koel-seven positions were generated for. The results of this ranking framework were combined with experimental measurements of virus fitness from these virus libraries. This was done in order to test whether combining historical statistical patterns with empirical measurements of virus fitness improved our ability to predict cluster transition substitutions. In order to know how well this method generalises to other cases, data from SI87 and BE92 should be censored for training purposes.

5.3.5 *Bumps plots*

Bumps plots are so-called because similar charts are used to display how rankings of rowing boats change throughout 'bumps' races.

Rankings of candidate substitutions from multiple antigenic clusters are visualised with bumps plots (e.g. Figure 5.5a). A zoomed in section of a bumps plot is shown in Figure 5.4. Rankings for each antigenic

cluster are arranged in separate columns ordered chronologically, left to right. Each row in each column shows a different candidate substitution. The same candidate substitutions are joined by lines if they appear in adjacent columns. This enables tracking how the rank of candidate substitutions changes in different clusters. Rankings change in different clusters for two reasons: (1) as amino acids in HA1 evolve, different substitutions are possible and (2) training data from preceding cluster transitions accumulate, which alters the relative probability associated with substitution features.

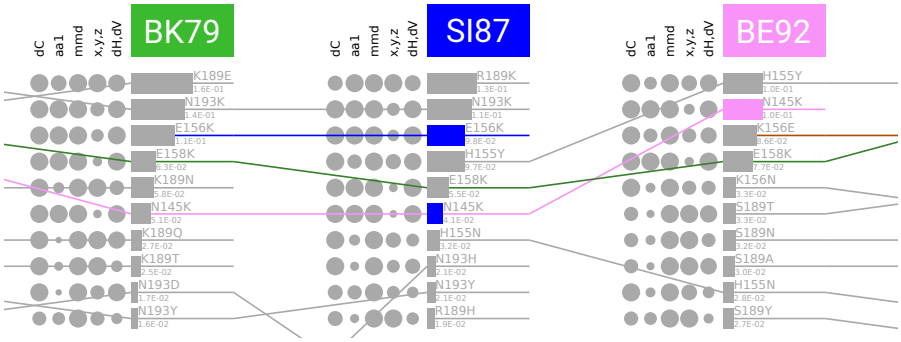


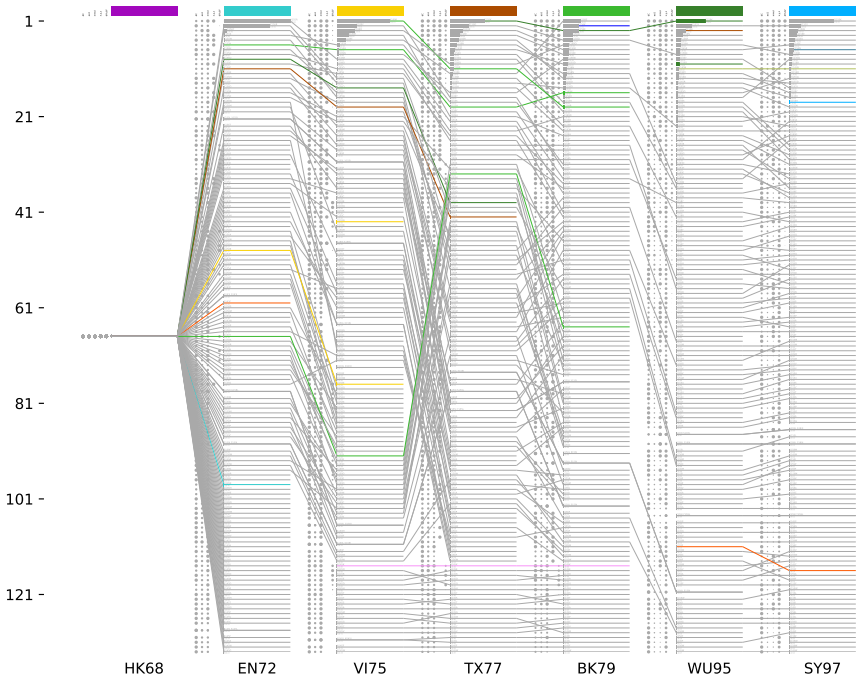
Figure 5.4: Bumps plot detail. This figure shows a zoomed in section of a bumps plot, in order to explain its elements. The top ten ranked substitutions from the BK79, SI87 and BE92 antigenic clusters in Figure 5.8. See Subsection 5.3.5 for a detailed description.

The thick main bar shows the value of h_s for a particular substitution. If the substitution is known to have caused a cluster transition in this antigenic cluster, it is coloured. Individual components of the h_s score from each feature are shown as circles to the left-hand side of the main bar. The area of the circle is proportional to the probability of the feature for that substitution, in that cluster. In the cluster that a substitution caused a transition from the main bar is coloured. For example Q156H is coloured light blue in the SY97 column because Q156H caused a cluster transition out of SY97 (Figure 5.5a). Substitutions with identical h_s are merged and shown in the same bar. This is particularly noticeable in the HK68 rankings (Figure 5.5a) for which there is no training data, and thus no basis to discriminate between any candidate substitutions.

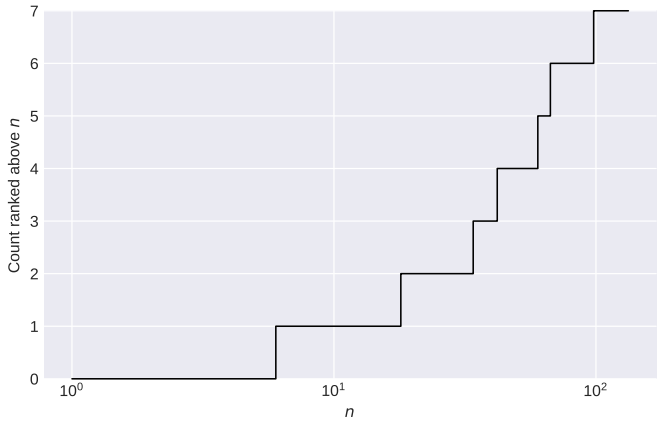
'Probability of the feature' refers to values of $P(x_{is})$ in Equation 5.3.

5.3.6 *Baseline rankings*

To serve as a baseline, I generated rankings for each antigenic cluster from HK68 to SY97 except BE89, SI87 and BE92 (Figure 5.5a). Training data for the EN72 ranking contains just a single substitution-the HK68–EN72 cluster transition substitution: T155Y. It is unsurprising therefore that the actual cluster transition from EN72–VI75, Q189K is ranked 98/133. By the final two cluster transitions, WU95–SY97 and SY97–FU02, enough training data has accumulated such that the



(a) Bumps plot (zoomable). See Subsection 5.3.5 for a detailed explanation.



(b) Number of cluster transition substitutions (y-axis) ranked above n (x-axis).

Figure 5.5: Baseline rankings. Baseline rankings were generated with no additional training data. Rankings in each column of (A) are based solely on training data from columns to their left. Candidate substitutions comprised only amino acid substitutions at Koel-seven positions, giving $19 \times 7 = 133$ candidates in each column. (B) shows that only a single substitution is ranked in the top ten positions.

rankings have improved modestly. The two substitutions responsible for the WU95–SY97 cluster transition are ranked 1^{st} and 10^{th}, and the substitution responsible for SY97–FU02 is ranked 18^{th}.

The rankings are summarised in a *count ranked above n* plot (Figure 5.5b) where the y-axis shows the number of cluster transition substitutions that are ranked above n, which is indicated on the x-axis. For cluster transitions caused by multiple substitutions, the mean rank of those substitutions is computed.

5.3.7 *Additional training data*

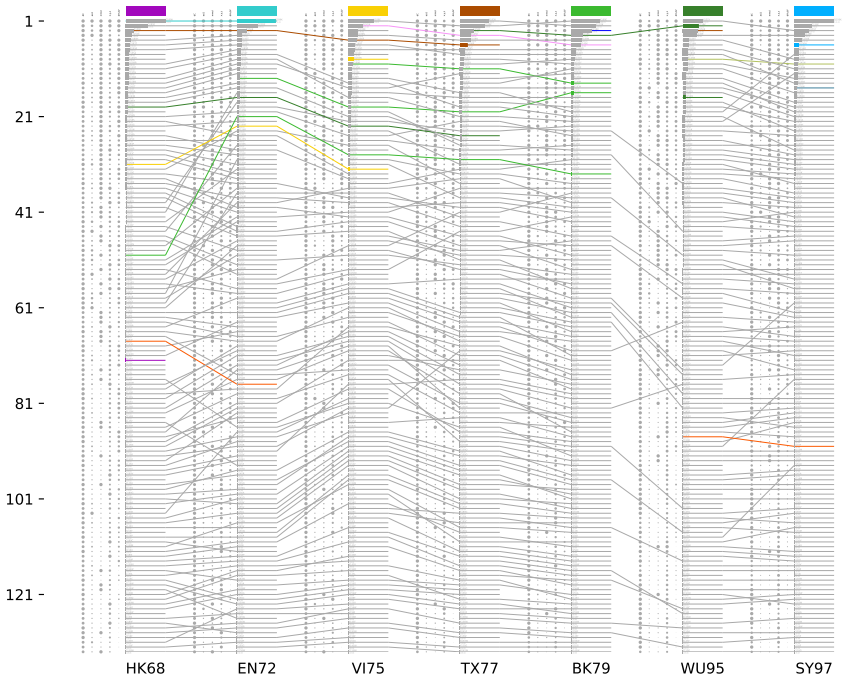
Next, I investigated whether using additional training data improved the rankings of human A(H3N2) cluster transition substitutions. Cluster transition substitutions are known in these non-'human A(H3N2)' influenza viruses: avian A(H5N1), equine A(H3N8), human A(H1N1), human B (Yamagata and Victoria lineages) and swine A(H3N2). These substitutions are collectively referred to as *additional training data*. They are visualised in network diagrams along with human A(H3N2) cluster transition substitutions in Figure 5.2.

Rankings were generated using the same basic setup as for the baseline rankings: the same set of candidate substitutions in the same set of clusters was used, and as before, only human A(H3N2) data from preceding cluster transitions was used. I first tested the impact of including all additional training data on the rankings (Figure 5.6). Seven out of eleven of the substitutions were ranked higher when all additional training data were included (mean increase of 37.2 places), the rank of one of the substitutions did not change, and three were ranked lower (mean decrease of 4.33 places). Three cluster transitions in the test set required multiple substitutions (VI75–TX77, BK79–SI87 and WU95–SY97). If for these cluster transitions, the mean rank of the multiple cluster transition substitutions is computed, then five out of seven were ranked higher when all additional training data are included (mean increase of 39.4 places) and two out of seven were ranked lower (mean decrease of 4.50 places). I conclude from this investigation that use of the additional training data improved the performance of the ranking framework.

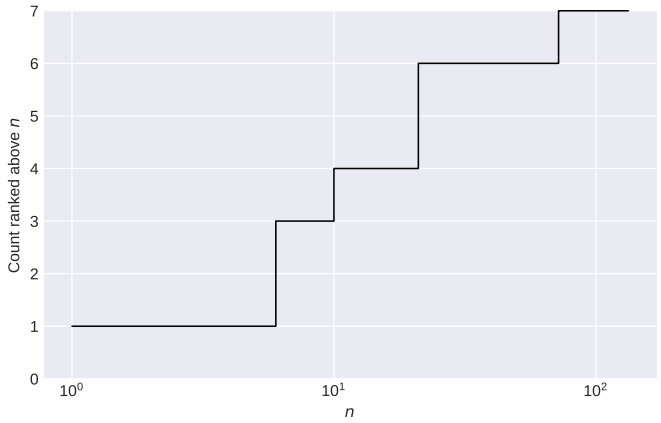
5.3.8 *Subsets of additional training data*

I then investigated whether the signal in each dataset from a non-'human A(H3N2)' influenza virus was beneficial or detrimental to the rankings by conducting two types of experiment. Each non-'human A(H3N2)' dataset was either used on its own as the additional training data, or all non-'human A(H3N2)' datasets except one were used as additional training data. As before, the same candidate substitutions in the same clusters were ranked; only the training data were modified.

Of the datasets used in isolation the avian and swine datasets performed best (Figure 5.7a). This may relate to the observation that these are the two largest datasets (Table 4.2). Ranks of cluster transition



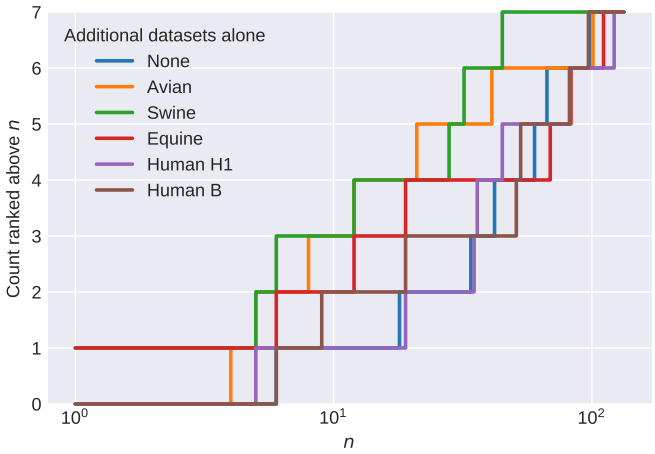
(a) Bumps plot (zoomable). See Subsection 5.3.5 for a detailed explanation.



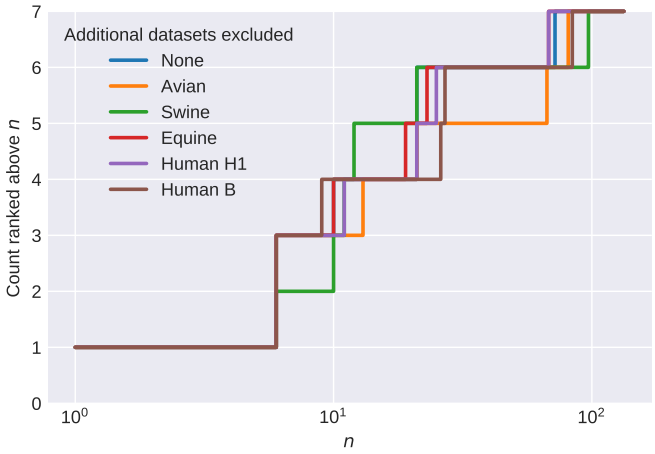
(b) Number of cluster transition substitutions (y-axis) ranked above n (x-axis).

Figure 5.6: Rankings including all additional training data. This is the same as Figure 5.5 except additional training data from cluster transition substitutions known in avian A(H5N1), equine A(H3N8), human A(H1N1), human B (Yamagata and Victoria lineages) and swine A(H3N2) influenza viruses were also included.

substitutions do not substantially decrease by using any individual dataset alone (Figure 5.7a). Similarly, ranks of cluster transition substitutions do not substantially increase by excluding any individual dataset (Figure 5.7b). I conclude that there is useful signal in all non-'human A(H3N2)' datasets, and therefore included all additional training datasets in future rankings.



(a) Each non-'human A(H3N2)' dataset used alone.



(b) All additional training data is used, and a single non-'human A(H3N2)' dataset is excluded. Here, None, means that no datasets are excluded.

Figure 5.7: Rankings using subsets of additional training data. The number of cluster transition substitutions (y-axis) ranked above n (x-axis) (a) including each non-'human A(H3N2)' training dataset alone and (b) including all additional training data and excluding one non-'human A(H3N2)' dataset.

Cluster transition	Substitution	Rank	Mean rank for cluster transition
SI87–BE92	E156K	3	3
SI87–BE89	N145K	6	6
BE92–WU95	N145K	2	2
FU02–CA04	K145N	8	8
CA04–WI05	S193F	85	85
WI05–PE09	K158N	11	6
	N189K	1	
PE09–SW13	F159S	101	101
PE09–HK14	F159Y	59	59

Table 5.2: Rank of cluster transition substitutions in test rankings.

5.3.9 *Testing*

A subset of human A(H3N2) cluster transition substitutions were purposefully withheld from the development of the framework in order to act as a blinded test set. The test set contained nine substitutions responsible for eight cluster transitions.

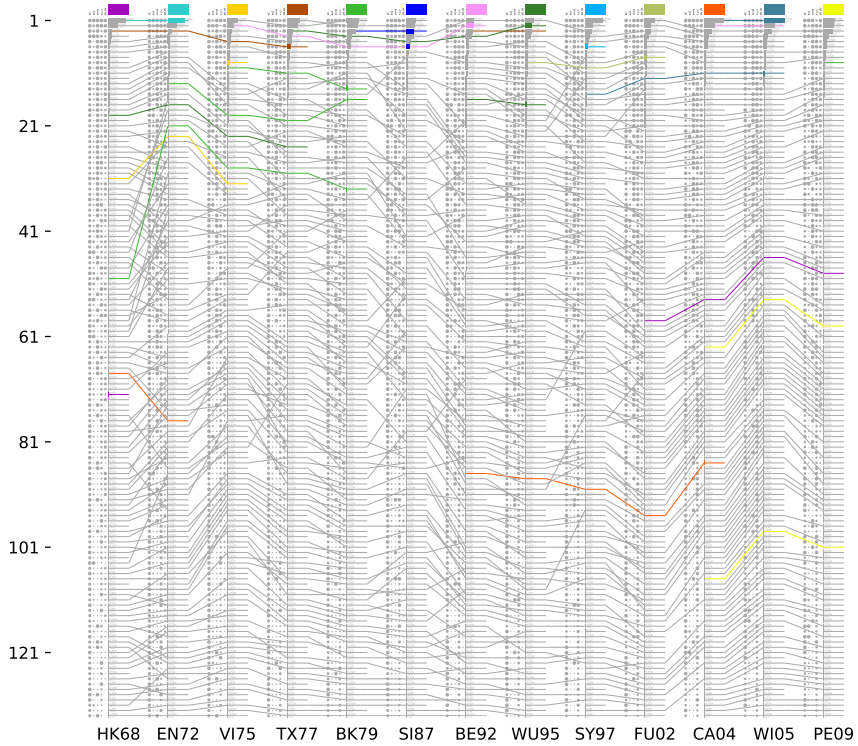
For testing, rankings were generated using the same strategy as before: candidate substitutions from each of the clusters were ranked according to h_s. As well as the additional, non-'human A(H3N2)' training data, human A(H3N2) cluster transition substitutions from preceding antigenic cluster transitions were used to train feature probability distributions to compute h_s.

Five out of nine substitutions were ranked in the top ten substitutions (Table 5.2, Figure 5.8). The test rankings are also visualised in an alternate representation which highlights the substitution properties (Figure 5.9).

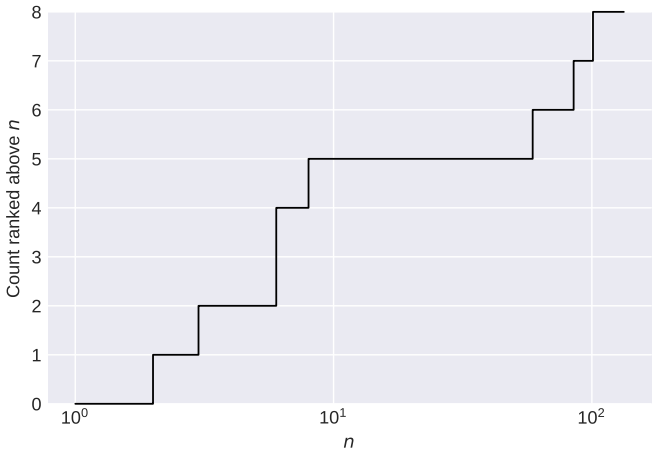
If rankings were generated at random, the probability of achieving a result at least this extreme is approximately:

$$\sum_{k=5}^{9} \binom{9}{k} \cdot \frac{10}{130}^{k} = 3.57 \times 10^{-4} \text{ (3 s.f.)}$$

This is approximate because for a cluster transition caused by two substitutions, the probability of observing both in the top ten at random is actually $\frac{10}{130} \cdot \frac{9}{129}$. This additional complexity is not incorporated in the calculation. I assume it would not impact the magnitude of the result which is dominated by the $\binom{9}{5} \cdot \frac{10}{130}^{5}$ term.



(a) Bumps plot (zoomable). See Subsection 5.3.5 for a detailed explanation.



(b) Number of cluster transition substitutions (y-axis) ranked above n (x-axis).

Figure 5.8: Test rankings. Cluster transitions used for testing are SI87–BE92, SI87–BE89, BE92–WU95, FU02–CA04, CA04–WI05, WI05–PE09, PE09–SW13 and PE09–HK14. For completeness, all earlier clusters that seeded another cluster are shown in the bumps plot.

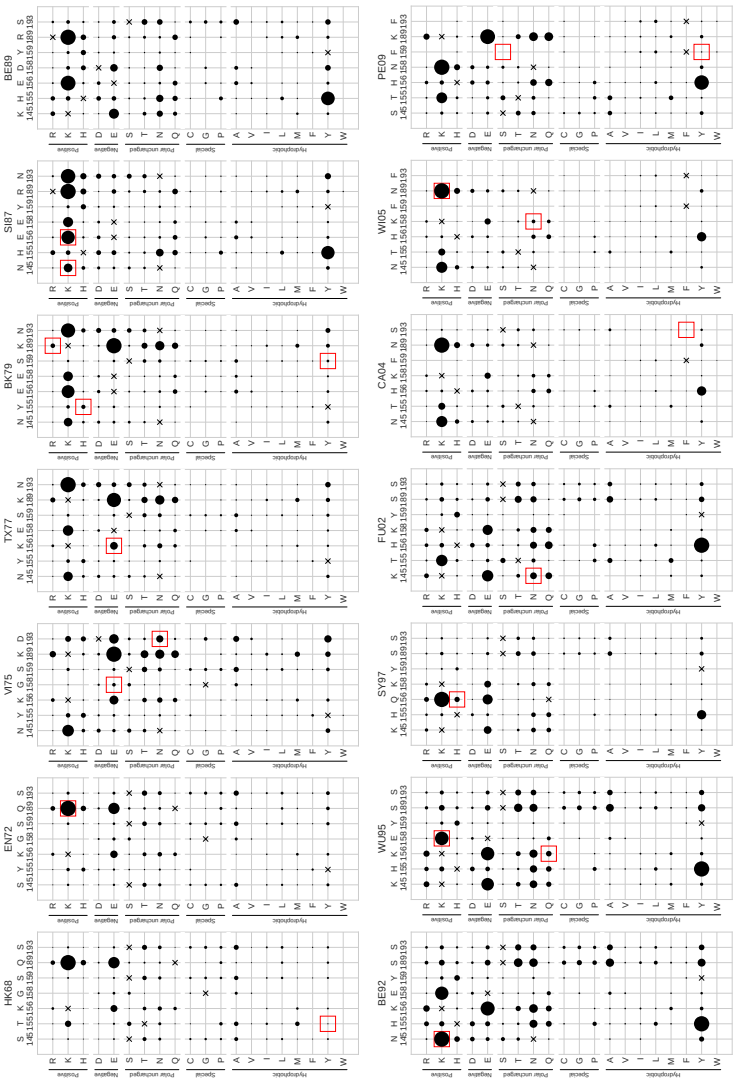


Figure 5.9: This is an alternate representation of the data visualised in Figure 5.8. Here, h_s scores for each substitution in a set of rankings are indicated by the area of a circle. Circles are arranged on a grid whereby the columns show the HA position of the substitution, and rows show the amino acid gained in the substitution. Crosses indicate the amino acids that are lost in the substitution. Amino acids are grouped loosely based on their properties. Red squares indicate the substitutions that caused cluster transitions. Circle areas are standardised such that the maximum size in each panel is the same. Only the top 100 substitutions in each set of rankings is shown.

5.3.10 *Extended candidate substitutions*

Up to this section, all rankings were conducted on candidate substitutions generated at the Koel-seven positions. Given the Koel et al. (2013) data which covers human seasonal influenza A(H3N2) viruses that circulated between 1968 and 2002, 14 cluster transition substitutions were observed at seven HA positions, and the same seven positions were repeatedly used. Furthermore, the substitutions identified in Chapter 4 that are responsible for the cluster transitions since 2002 also only involved these Koel-seven HA positions. However, the possibility of additional HA positions harbouring cluster transition substitutions cannot be ruled out. A simple model may shed light on how likely this is.

The process that selects cluster transition substitution positions can be modelled similarly to rolling a die. The number of faces the die has represents the number of HA positions that can harbour cluster transition substitutions. We do not know how many positions cluster transition substitutions can occur at, but in the data so far we have seen seven positions. In the die analogy, we do not know how many faces the die has, but we have rolled it several times and so far have only observed seven faces. The question becomes: how likely is it that the die actually only has seven faces, if we've seen seven faces? Furthermore, how does this likelihood change as we see more data?

After 40 rolls only seeing seven faces, it is overwhelmingly likely that there are only seven faces (Figure 5.10). After 14 rolls only seeing seven faces, it is actually more likely that the die has eight faces, although the probability that the die has seven or nine faces are only slightly lower.

This model assumes an unbiased die, which may not be realistic. In the Koel et al. (2013) data position 156 is observed four times, whereas position 159 is observed only once.

This model is extremely reductionist and therefore the exact probabilities should not be used directly in additional calculations. I use it to highlight that even in an extremely simple model, we should attribute a non-zero probability to the possibility to there being additional positions that can harbour cluster transitions.

In this section, I relax the assumption that future cluster transitions can only be caused by amino acid substitutions at a Koel-seven positions. I achieve this by extending the number of HA positions that candidate substitutions for rankings are generated at. Instead of only Koel-seven positions, I included any solvent exposed position within 35 Å of a Koel-seven position in the A/Hong Kong/1/1968 (H3N2) crystal structure (PDB ID: 4FNK (Ekiert et al. 2012)). The 116 positions that meet these criteria were: 57, 58, 59, 60, 62, 63, 65, 74, 75, 77, 78, 80, 81, 82, 83, 85, 91, 92, 93, 94, 95, 96, 100, 101, 103, 104, 105, 106, 114, 119, 121, 122, 123, 124, 125, 126, 128, 129, 131, 132, 133, 134, 135, 136, 137, 140, 141, 142, 143, 144, 145, 146, 149, 150, 155, 156, 157, 158, 159, 160, 162, 163, 165, 166, 167, 168, 169, 171, 172, 173, 174, 175, 186, 187, 188, 189, 190, 192, 193, 194, 196, 197, 198, 199, 201, 207, 208, 209, 210,

A(H3N2) specific numbering was used throughout (Burke and Smith 2014).

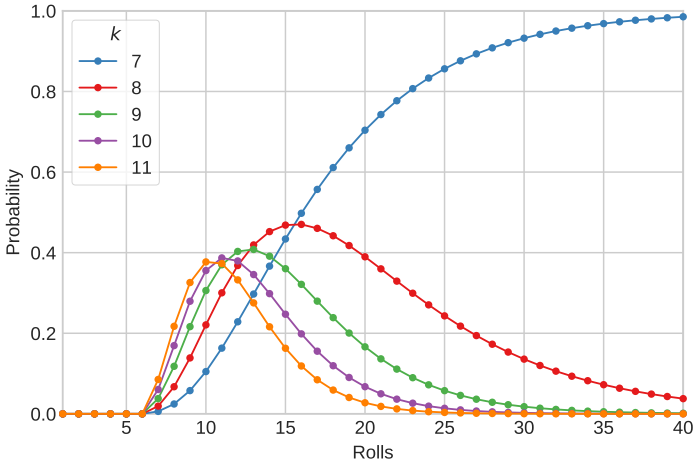


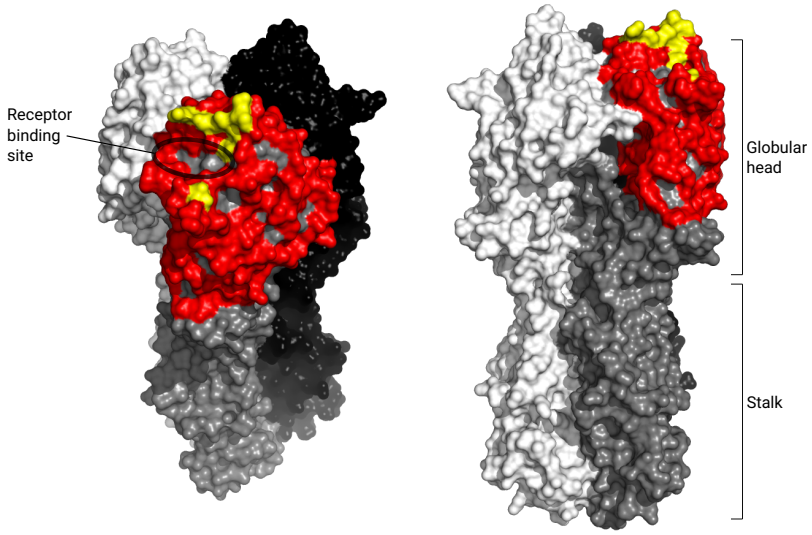
Figure 5.10: Probability a die has k sides having observed seven. If a fair die has an unknown number of faces, repeatedly rolling it and observing the faces can provide evidence on how many total faces there are. Here, the y-axis shows the probability that a fair die with k sides generated data containing exactly seven faces for a varying number of rolls (x-axis).

211, 212, 213, 214, 219, 222, 223, 224, 225, 226, 227, 233, 234, 236, 238, 239, 240, 241, 242, 248, 255, 259, 261, 269, 271, 273, 274. See Figure 5.11 for these positions displayed on the A/Honk Kong/1/1968 (H3N2) crystal structure.

Rankings were generated using candidate substitutions at all 116 of these HA positions. Each ranking contains $116 \times 19 = 2,204$ substitutions. See Figure 5.12a for a summary of the top 100 positions of each of these rankings, and Figure 5.12b for a summary of the positions of known cluster transition substitutions in these rankings. Similarly to the test rankings, in over half of cases, cluster transition substitutions were ranked in the top ten substitutions.

5.4 DISCUSSION

The ability to rank cluster transition substitutions as effectively as demonstrated here stems from their pronounced biophysical patterns. These patterns appear to be a general feature of amino acid substitutions that cause major antigenic change in diverse influenza viruses (Figure 5.2). This observation suggests that the patterns derive from mechanisms that alter antigen-antibody interactions that are not specific to a particular influenza virus subtype or host. Large biophysical changes, such as changes in side chain volume and formal charge, may disrupt epitope-paratope complementarity regardless of host or subtype. Similarly, the maintenance of residue hydropathy suggests that maintaining solvent interactions is important regardless of host or subtype. A pragmatic benefit of this similarity across different

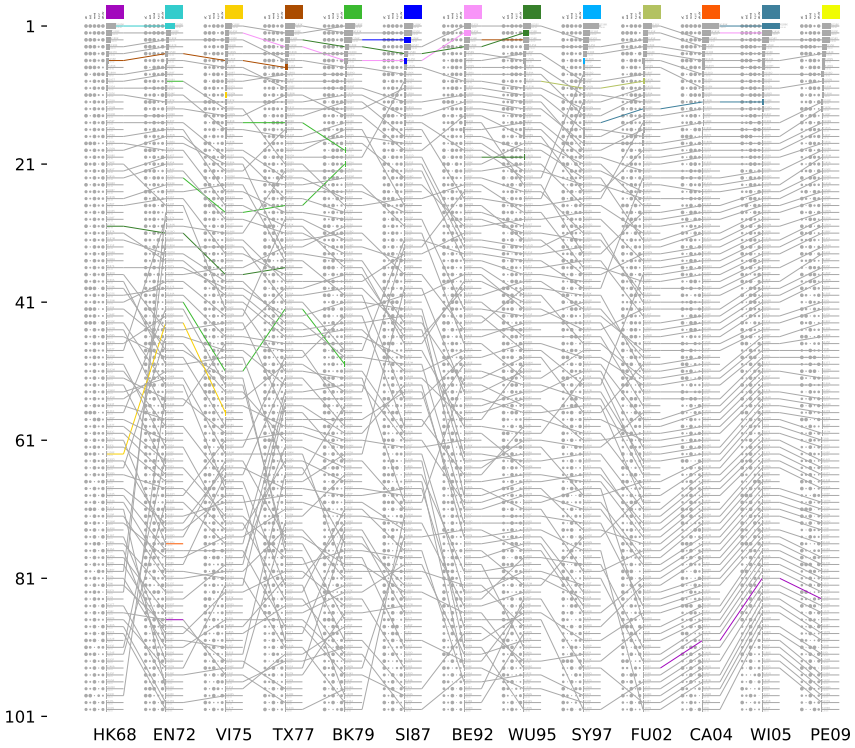


(a) (b)

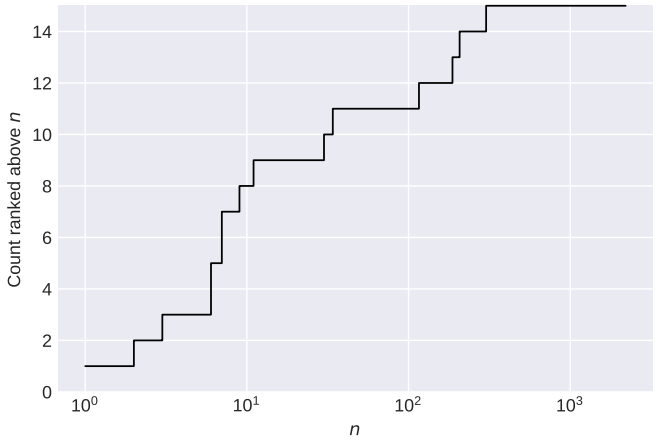
Figure 5.11: HA positions of extended candidate substitutions. (a) and (b) are different rotations of the same crystal structure. The three HA monomers are shown in black, grey and white. On the grey monomer Koel-seven positions are yellow and the extended positions are red. This figure was rendered using PyMOL (Schrödinger LLC 2015) on the A/Hong Kong/1/1968 (H3N2) crystal structure, PDB ID: 4FNK (Ekiert et al. 2012).

hosts and subtypes is to increase the amount of training data which improves retrospective rankings (Figure 5.8). This additional training data probably also improves how generalisable into the future this ranking framework will be.

The ranking framework is not based on a mechanistic understanding of the processes which govern influenza virus antigenic evolution. Instead, patterns are observed in training data, abstracted as probability distributions, and naïvely applied to rank candidate substitutions. The patterns observed are consistent with our expectations. For instance, it is unsurprising that substitutions that induce large changes in physicochemical properties disrupt epiotope-paratope complimentary. However, although the rankings perform remarkably well, they do not allow us to understand why the patterns seem to apply in some cases, such as the many instances of involving lysine, but not in others, such as T155Y, responsible for the HK68–EN72 cluster transition, which is ranked 72nd in the test rankings (Figure 5.8). A more detailed mechanistic understanding of the processes that govern the antigenic evolution of influenza viruses may allow us to understand why some cluster transition substitutions follow historical patterns and others do not.



(a) Bumps plot (zoomable). See Subsection 5.3.5 for a detailed explanation.



(b) Number of cluster transition substitutions (y-axis) ranked above n (x-axis).

Figure 5.12: Wider candidate substitution rankings. These rankings show the top 100 out of 2,204 candidate substitutions generated for an extended set of solvent exposed HA positions around the globular head.

The rankings presented in this chapter are substantially better than if substitutions were ranked at random. This is scientifically interesting, but there are plenty of cases where the correct substitution is not ranked first. Superficially, this may seem to prevent this work from being useful to inform vaccination strategies. If there was even a very

low chance that substitution ranked first was not going to cause the next cluster transition substitution, would it be ethical to deploy a strain with this substitution in a vaccine?

As outlined in the introduction to this chapter, an immunity management strategy does not require full confidence in predictions of future antigenic variants. Thanks to the back-boost, an antigenically advanced strain generates the same recall of prior immunity as using a matched vaccine. Therefore, if the predicted antigenic variant does not circulate, protection conferred by an advanced vaccine against non-drifted strains may be the same as a matched vaccine.

In the next and final chapter I introduce experimental measurements of intrinsic and extrinsic fitness that have been carried out on libraries of influenza virus mutants containing all possible single substitution mutants at the Koel-seven positions. Combining insights from the rankings presented in this chapter and the fitness assessments yields further improved predictions.

6

PREDICTING THE ANTIGENIC EVOLUTION OF SEASONAL INFLUENZA VIRUSES

CHAPTER OUTLINE In this final chapter I summarise the thesis and discuss it in the context of a wider body of work being carried out by other members of my group and our collaborators.

6.1 SUMMARY

This thesis outlines the predictability of substitutions that cause major antigenic change, and quantifies the impact that solving antigenic mismatch would have on VE. In Chapter 3 I conducted the first detailed analysis on the impact of antigenic mismatch on VE using measurements of antigenic mismatch derived from antigenic maps. The results suggest that there is a threshold value of approximately 3.5 AU mismatch above which VE begins to drop steeply. Questions have been raised about the legitimacy of VE measured by the TND. The observation of a clear pattern between seasons with higher antigenic mismatch and lower VE broadly validates the VE signal measured by TND. Antigenic coordinates of some viruses in Chapter 3 were derived from BLUP predictions, which I introduce in Chapter 2 as part of a novel application of LMMs developed in the field of quantitative genetics.

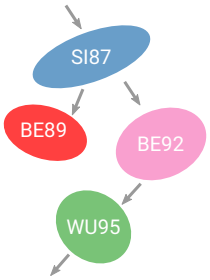
In Chapter 4 I reviewed current knowledge on substitutions that cause major antigenic change in diverse influenza viruses. I went on to identify additional substitutions responsible for cluster transitions in human A(H3N2) viruses using a combination of natural experiments and statistical tests in the LMM framework. The substitutions I identified continued pronounced patterns in other known cluster transition substitutions.

In Chapter 5 I demonstrated that cluster transition substitutions are highly non-random, allowing me to develop a framework to rank candidate substitutions by their similarity to cluster transition substitutions. I also outlined a novel immunity management vaccination strategy which proposes using antigenically advanced vaccines to build immunity in the population ahead of the antigenic evolution of the virus. I then summarised data from Fonville et al. (2014) showing that antigenically advanced vaccines generate the same recall of prior immunity as matched vaccines. Due to the back-boost, antigenically advanced vaccines may not be inferior to antigenically matched vaccines, even if drift variants do not circulate. This implies that complete confidence in the identity of future variants is not required for this work to have translational impact.

Of course, higher confidence predictions are desirable, and depend on honing our understanding of the processes which govern the antigenic evolution of these viruses. The work presented in this thesis is one component of a larger project to predict the antigenic evolution of seasonal influenza viruses. This project is a collaboration between Prof. Smith's group at the University of Cambridge, U.K. and virologists at the University of Wisconsin, Madison, U.S.A.; Erasmus Medical Centre, Rotterdam, Netherlands; and the Peter Doherty Institute for Infection and Immunity, Melbourne, Australia. I will now outline this project and the role of this thesis in it.

6.2 ALL SUBSTITUTION EXPERIMENT

Koel et al. (2013) showed that in most cases cluster transitions are caused by single amino acid substitutions at only seven HA positions in human A(H3N2) influenza viruses. This pattern continued to hold for the later cluster transition substitutions I identify in Chapter 4. These findings imply that only $19 \times 7 = 133$ single amino acid variants would need to be generated to conduct a screen containing the future cluster transition substitution.



Schematic view of clusters involved in the all substitution experiment.

Such mutant screens were conducted for two retrospective cases. Shufang Fan at the University of Wisconsin, Madison, U.S.A. generated all single amino acid variants at Koel-seven positions in a strain from the SI87 antigenic cluster, and separately in a strain from the BE92 cluster. SI87 was selected because it is the source of a bifurcation in the otherwise stepwise progression of clusters (Figure 1.4). Using SI87 therefore enabled investigating two antigenic cluster transitions with a single set of mutants.

Antigenic phenotypes of these viruses were measured by Malet Aban at the Peter Doherty Institute for Infection and Immunity, Melbourne, Australia. A large HI assay was conducted using a panel of 16 antisera known to accurately coordinate viruses in this region of antigenic space. Sarah James and I, at the Centre for Pathogen Evolution, Department of Zoology, University of Cambridge, U.K., used the titrations to generate an antigenic map (Figure 6.1).

The maps comprise an exhaustive exploration of the antigenic space accessible by single substitutions at Koel-seven positions. Surprisingly, the majority of mutants did not escape the cluster of their parent virus. From the SI87 virus, of the mutants that did escape the cluster, only two directions are observed; either towards the subsequent BE89 cluster, or towards the subsequent BE92 cluster. Similarly, from the BE92 strain, the majority of mutants do not escape the cluster, and those that do, radiate towards the subsequent cluster, WU95. The BE92 K156E mutant does not head towards WU95, but towards SI87. K156E is the reversion of E156K which caused the SI87–BE92 cluster transition. Despite generating all possible single amino acid substitution mutants

at Koel-seven positions, most substitutions show low antigenic impact. Substitutions that do have an antigenic impact appear to be restricted to certain areas of antigenic space.

Generally speaking, predictions are better when they combine data from multiple different sources. This is the case here. Figure 6.1B–G show smaller regions of the all substitution antigenic map. In Figure 6.1F and G the area of each mutant is proportional to h_s. Many substitutions with high h_s are not antigenically novel. Therefore, although they are ranked highly based on similarity to past cluster transition substitutions, we would not expect them to cause a cluster transition because they do not escape prior immunity. Conversely, antigenically novel mutants that are dissimilar to historical cluster transition substitutions on the basis of h_s can also be discounted. N145K caused the SI87–BE89 cluster transition. Of the mutants that radiate downwards from SI87 N145K has the second highest h_s. Similarly, of mutants that radiate to the right from SI87, E156K has the highest h_s and caused the SI87–BE92 cluster transition. Finally, of BE92 mutants that radiate away from prior clusters, N145K, which was selected in nature, also has the highest value of h_s. Therefore, combining h_s with experimental measurements of antigenic phenotypes yields good retrospective predictions of cluster transition substitutions.

h_s is described in Chapter 5 and is used to rank substitutions.

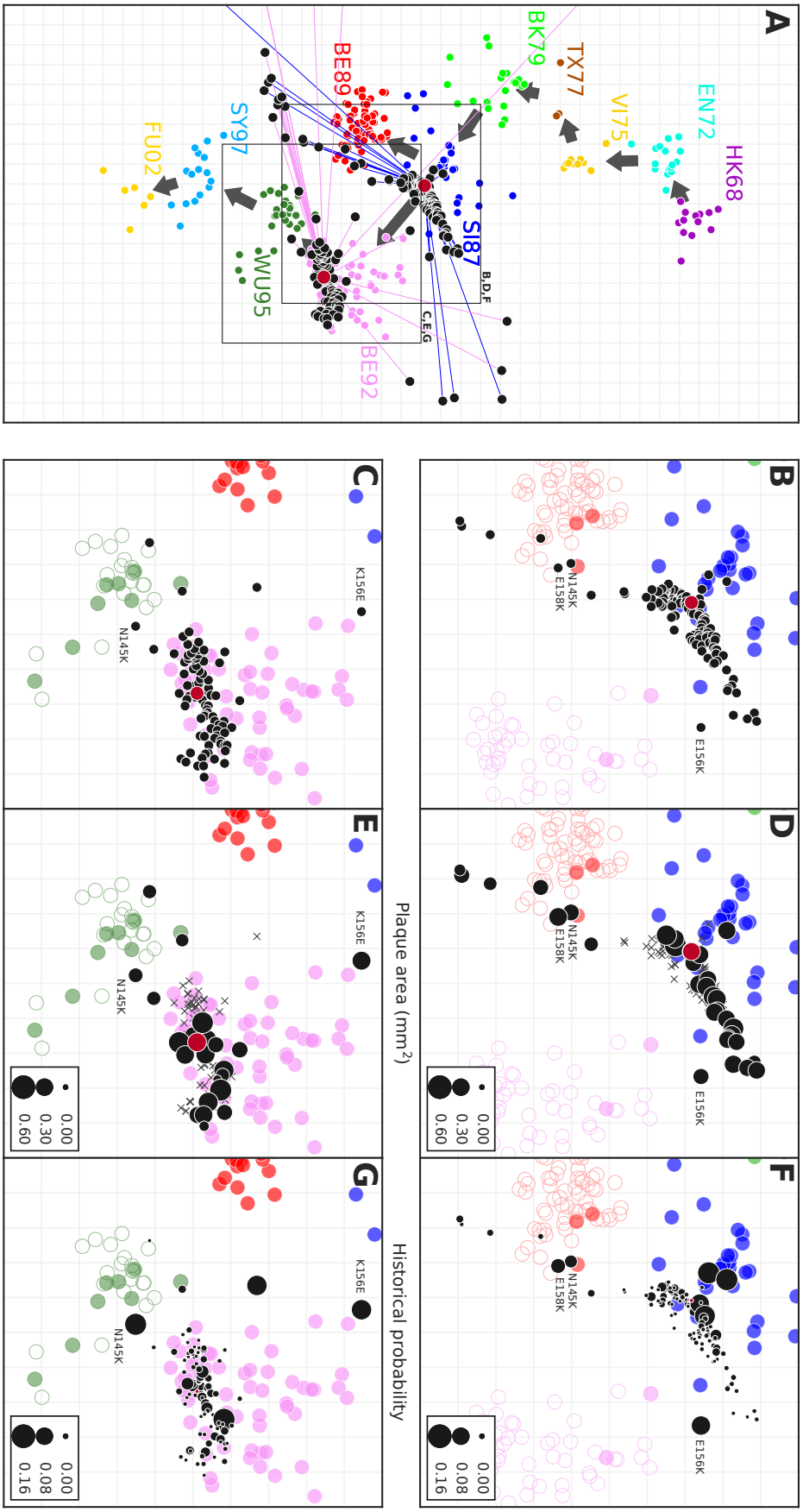


Figure 6.1: Antigenic map of single substitution mutants. (Continued on next page.)

Figure 6.1: (Previous page). (A) Antigenic coordinates of the single substitution mutant viruses from the Sichuan 1987 (SI87) and Beijing 1992 (BE92) antigenic clusters are shown in black. Large dark red marks show the wild type viruses (SI87, A/H3N2/Hong Kong/1/1989; BE92, A/H3N2/Hong Kong/56/1994), and are connected to their mutants by blue (SI87) and pink (BE92) lines. Three mutants had no detectable titres and are therefore not shown on the map. Surrounding surveillance strains are also shown and coloured according to antigenic cluster Smith et al. (2004). (B, C) Zoomed-in maps show the majority of the BE92 and SI87 all substitution mutants separately. Strains in clusters that circulated directly after SI87 (BE89 and BE92), and BE92 (WU95) are shown as empty circles in the SI87 and BE92 panels, respectively. Strains that circulated in the first season of the subsequent clusters are filled in lightly. (D, E) The area of each mutant is proportional to the median area of plaques in the plaque assay. Mutants not measured in the plaque assay are shown with crosses. (F, G) The area of each mutant is proportional to historical probability.

6.3 FITNESS COST OF ANTIGENIC CHANGE

Other data may also help improve predictions and our understanding of processes that govern the antigenic evolution of seasonal influenza viruses. Antigenic mapping measured the extrinsic fitness of the all substitution mutant viruses, i.e. the degree to which a mutant escapes host immunity built up to older clusters. Intrinsic fitness refers to all other components of virus replicative fitness (Subsection 1.4.1). Ramona Mögling at the Erasmus Medical Centre, Rotterdam, Netherlands, developed a specialised plaque assay to measure small differences in replicative fitness between viruses, and used this assay on a selection of the all substitution mutants. Plaque size is shown in Figure 6.1 D and E as the area of each mutant.

Viruses which replicate faster produce larger plaques.

Here, I am only showing plaque assay data. But, plaque growth correlates well with other assays used to measure intrinsic fitness including a competition assay and a next-generation sequencing FRA assay currently under development in my lab in collaboration with virologists at the Erasmus Medical Centre.

Plaque size appears to be less discriminatory than h_s for predicting cluster transition substitutions. There are several cases where antigenically novel strains have plaque sizes that are larger than the strain that was selected in nature. However, a striking pattern emerges from Figure 6.1 D and E: plaque sizes reduce with increasing distance from the parent virus. This is visualised explicitly in Figure 6.2. There appears to be a replicative fitness threshold above which mutants do not exist, and this threshold appears to decrease with increasing antigenic distance. More simply, no mutants exist at high antigenic distance and at high replicative fitness. Substitutions that produce gains in extrinsic fitness appear to have an intrinsic fitness cost.

This relationship is expected given the position of cluster transition substitutions on the periphery of the HA RBS. Any substitution in this region is likely to impact receptor binding in some way, and therefore impact replicative fitness. The observed relationship between intrinsic and extrinsic fitness was actually hypothesised by Koel et al. (2013) based on the HA positions of cluster transition substitutions. It also has implications for our understanding of processes that govern the antigenic evolution of seasonal influenza viruses.

In Chapter 1 I introduced the fitness exchange hypothesis. To reiterate, the fitness exchange hypothesis explains cluster transition delay as the time during which extrinsic fitness benefits of cluster transition substitutions are outweighed by their intrinsic fitness costs. Therefore, cluster transitions occur when sufficient host immunity has developed in the population such that the intrinsic fitness loss is offset by the extrinsic fitness gain. The empirical relationship between antigenic escape and replicative fitness demonstrated in Figure 6.2 is entirely consistent with this mechanism.

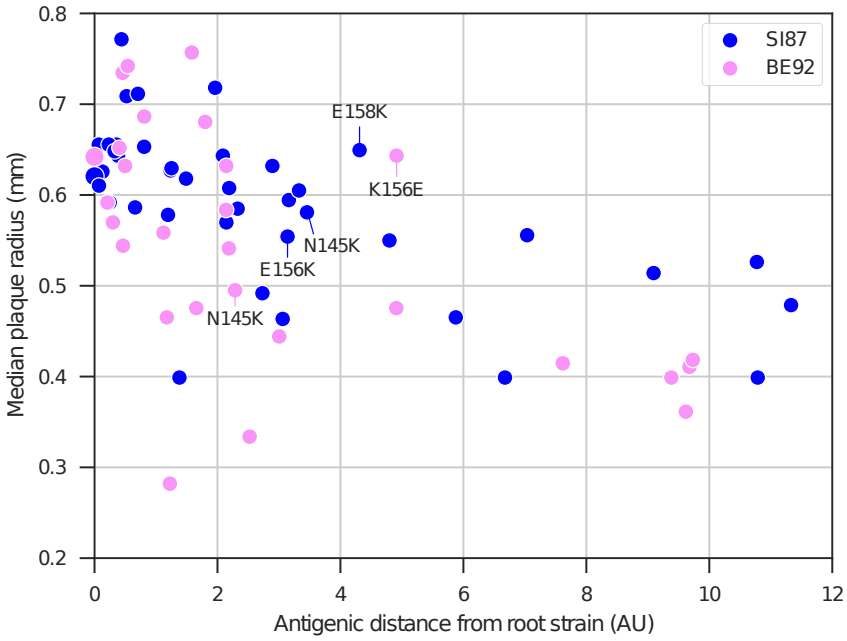


Figure 6.2: In-vitro fitness cost of antigenic change. Median radius of plaques from plaque assay versus antigenic distance to wild type for SI87 (blue) and BE92 (pink) mutants. N145K caused the cluster transition from the SI87 to BE89, E156K from SI87 to BE92, and N145K from BE92 to WU95. The SI87 and BE92 wild type viruses are the two larger dots with an antigenic distance of 0 AU.

Despite finding evidence consistent with the fitness exchange hypothesis based on antigenic distance and replicative fitness, these measurements alone do not yield predictions as good as when h_s is also included. h_s must distil important factors regarding why particular substitutions cause cluster transitions in nature. In some cases h_s features have plausible mechanisms. For example, only a single cluster transition substitution is known that required more than one nucleotide mutation, which suggests that the rate at which substitutions are sampled during an infection is an important factor. Plausible mechanisms underlying other cluster transition substitution features are outlined in Section 5.4. In other cases, patterns are stark, but mechanisms are less obvious. Why do so many cluster transition substitutions involve either the loss or a gain of a lysine residue? Regardless, patterns abstracted in h_s provide orthogonal signal to extrinsic and intrinsic fitness assays. Strikingly, it is the combination of computational and wet lab approaches that provides the most powerful analysis.

6.4 OUTLOOK

The all substitution experiment was conducted on two viruses from different antigenic clusters. Although this required substantial work in four laboratories, we still 'only' have data relevant to three cluster transitions. We are prospectively testing these methods on currently circulating seasonal influenza viruses and are planning additional retrospective tests. Of particular interest are cluster transitions where the rankings perform poorly (e.g. HK68 and CA04, Figure 5.8). In the case where two nucleotide mutations were required, did no single mutation substitutions have suitable intrinsic and extrinsic fitness? Developing higher throughput intrinsic and extrinsic fitness assays is crucial for generating the data required to refine our understanding of this system. This is also under active development in our consortium.

Ultimately, phenotypic effects of substitutions are mediated by the structural biology of virus and host proteins. Changes in biophysical properties of residues may disturb antigen-antibody interactions, or alter complementarity between the RBS and sialic acid receptors. We have a poor understanding of these processes at a structural level, and how they impact virus fitness. This understanding would enable more sophisticated screening of future antigenic variants, and more powerful experiments to test processes that govern the evolution of these viruses. Computational structural biology provides attractive methods for addressing these questions.

Part I

APPENDIX

A

MAP SANITISING

In this appendix I introduce the concept of map sanitising. This is a novel filtering method based on BLUP, which can be used to display only the antigenic variation in a map which is associated with molecular change, thereby 'removing' some map error.

The Smith et al. (2004) antigenic map is qualitatively different to the Russell et al. (2008b) and Li et al. (2016) antigenic maps. In the Smith et al. (2004) map different antigenic variants are separated into distinct separate clusters. Conversely, there is less obvious spatial separation between the major antigenic variants in the Russell et al. (2008b) and Li et al. (2016) maps. Close inspection of the Russell et al. (2008b) and Li et al. (2016) maps reveals denser regions, delimited and labelled in Figure 2.4, which seem to roughly correspond to the distinct clusters in the Smith et al. (2004) map. Several hypotheses may explain these differences.

Firstly, the Russell et al. (2008b) and Li et al. (2016) maps use assay data derived from the GISRS which may contain additional sources of error compared to the Smith et al. (2004) dataset. The surveillance data are derived from hundreds of individual assays, which will introduce additional variables such as different technicians, reagent stocks and cell types used to passage viruses. Furthermore, due to the high throughput of the GISRS, outlier measurements are not usually investigated or repeated, and very rarely are additional titrations conducted to improve map coordination. In contrast, the Smith et al. (2004) titrations were conducted by a single technician in relatively few distinct assays, outlier titrations were double-checked, and additional titrations were conducted to improve the triangulation in specific regions of the map.

Secondly, viruses isolated when the Russell et al. (2008b) dataset was collected showed variable binding preferences to red blood cells. This resulted in a change in the type of red blood cells recommended for use in HI assays (Subsection 1.3.3). Altered binding preferences to red blood cells may explain additional variance in HI assay measurements during this period.

Thirdly, the Russell et al. (2008b) and Li et al. (2016) maps have higher temporal sampling density than the Smith et al. (2004) map. Smith et al. (2004) sampled 7.8 antigens per year; Russell et al. (2008b), 283.2 antigens per year; and Li et al. (2016), 774.3 antigens per year. This may have caused fewer minor antigenic variants to be sampled in the Smith et al. (2004) dataset, causing the major antigenic variants to appear more distinct.

These counts use the number of antigens that had their HA1 sequenced and were also titrated for inclusion in the map.

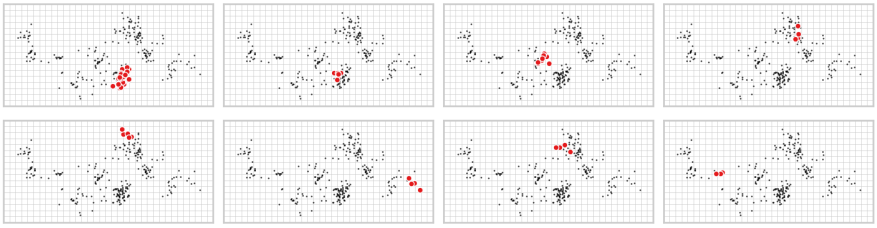
One way to measure the amount of error in an antigenic map is to consider groups of genetically identical strains. In a system with no experimental error, if two genetically identical strains were titrated against the same panel of antisera and included in an antigenic map, then you may expect them to have identical antigenic coordinates. This is not observed: strains with identical HA1 sequences have a median intra-group distance of 1.21 AU in the Smith et al. (2004) map; 1.83 AU in the Russell et al. (2008b) map; and 1.15 AU in the Li et al. (2016) map (Figure A.1d).

One explanation for this may be that variation in other segments causes the antigenic differences. Indeed, although HA1 is the main determinant of HI activity, NAs of recently circulating A(H3N2) influenza viruses have haemagglutinating activity (Mögling et al. 2017). HA and NA are the two main glycoproteins on the surface of influenza virus particles (Figure 1.1), so it is unlikely that other proteins would influence HI activity. To test whether genetic variation in NA may cause antigenic variation, I visualised groups of strains with identical HA1 and NA sequences in the Li et al. (2016) dataset (Figure A.2). Strains with genetically identical HA1 and NA have a median intra-group distance of 1.16 AU, which is similar to the value of 1.14 AU when just considering groups with identical HA1. Therefore, NA variability does not explain this antigenic variation.

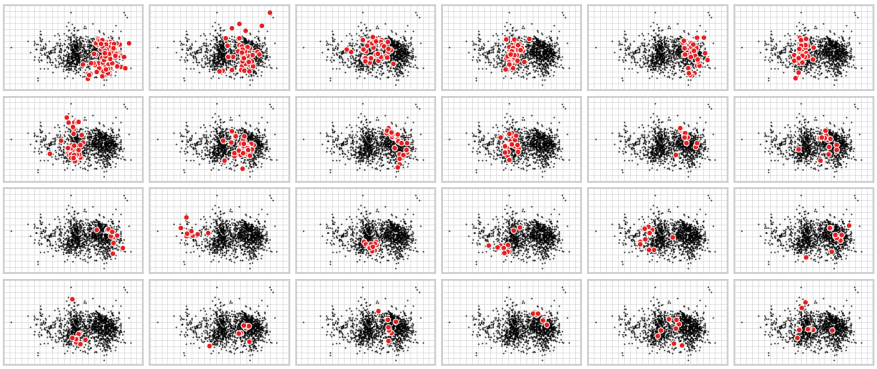
Another explanation is that antigenic variation in genetically identical strains is caused by measurement error in the assay, and poor coordination of antigens. This prompted the development of a framework to 'sanitise' antigenic maps based on the genetic sequences of strains.

In map sanitising, a model which maps strain genotypes to antigenic phenotypes is trained, and subsequently used to predict the antigenic phenotypes of each strain in a dataset. Here, I use BLUP, but map sanitising could be done by any model which maps genotype to antigenic phenotype. Concretely, for map sanitising using BLUP, a dataset containing antigenic coordinates and AAPs are used as the in-sample dataset, and parameters in Equation 2.14 are learnt, as outlined in Section 2.3. Then, the same dataset is treated as the out-of-sample dataset, for which phenotype predictions are made.

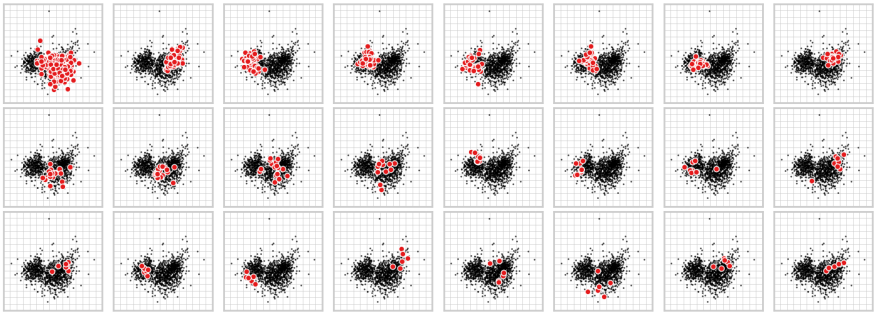
I applied map sanitising to the Smith et al. (2004), Russell et al. (2008b) and Li et al. (2016) datasets (Figure A.3). The Russell et al. (2008b) and Li et al. (2016) maps look qualitatively more like the original Smith et al. (2004) map. The size of the overall footprint of each of these maps is highly reduced, and antigenic variants in different regions of the map are more distinct, similar to the clusters in the Smith et al. (2004) map. These differences are also observed comparing the original Smith et al. (2004) map to its sanitised version, although the differences are less extreme.



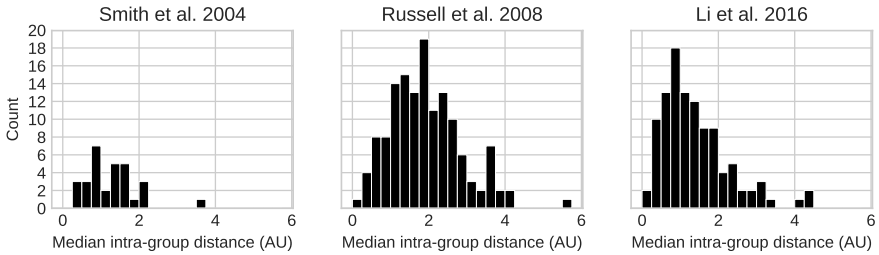
(a) Smith et al. (2004). The largest eight out of 30 groups.



(b) Russell et al. (2008b). The largest 24 out of 140 groups.

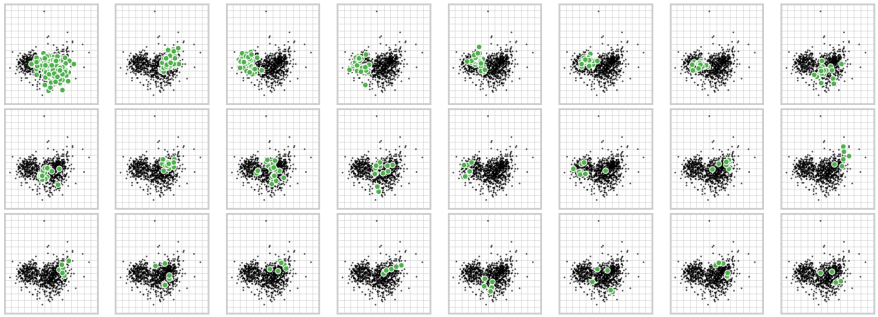


(c) Li et al. (2016). The largest 24 out of 114 groups.

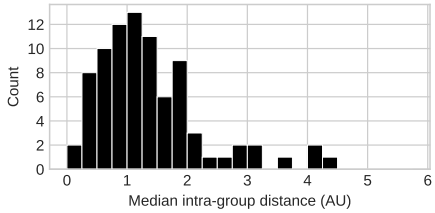


(d) Summary of median intra-group distance

Figure A.1: Antigenic positions of strains with identical HA1 sequences. (a-c) Antigenic maps highlighting a group of strains with identical HA1. Panels are ordered such that groups decrease in size from left to right, top to bottom. (d) Distributions of median intra-group distances for each dataset-.



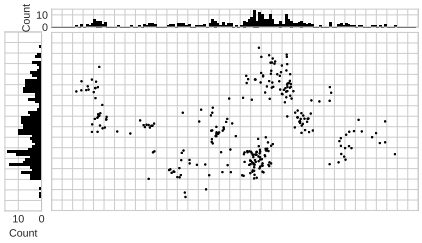
(a) Groups of strains with identical HA1 and NA sequences.



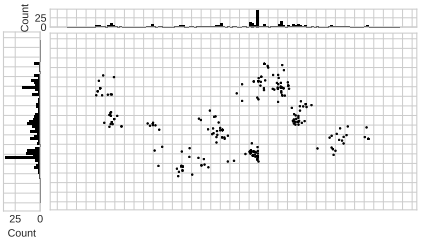
(b) Median intra-group distances.

Figure A.2: Antigenic positions of strains with identical HA1 and NA sequences in the Li et al. (2016) antigenic map. (a) Each panel shows an antigenic map with a group of strains with identical HA1 and NA sequences highlighted in green. Panels are ordered such that groups decrease in size from left to right, top to bottom. 84 groups were identified in total, only the largest 24 groups are shown. (b) The distribution of median intra-group distances.

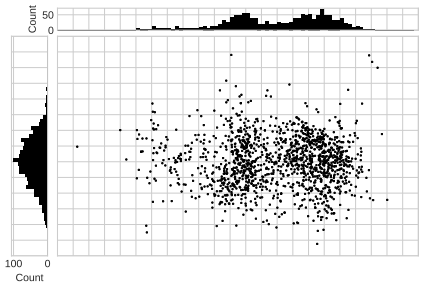
The smaller footprint of the sanitised maps suggest that the larger footprint observed in the original maps do not have a strong underlying molecular basis. This supports the hypothesis that these large footprints are caused by assay error rather than the Russell et al. (2008b) and Li et al. (2016) datasets containing many more additional antigenic variants due to more intensive sampling. One way to test this further would be to create maps containing antigens covered by the Russell et al. (2008b) and Li et al. (2016) datasets, but without the methodological drawbacks implicit in the GISRS data: careful selection of a panel of antisera that coordinate viruses well in antigenic space; repeating outlier titrations; individual technicians; and individual batches of reagents. These projects are ongoing in my lab with wet-lab collaborators. If such maps have smaller footprints, and are more like the sanitised versions of the GISRS maps than the original GISRS maps, then sanitising may become a useful post processing analysis to conduct when using GISRS data or any large dataset that may contain substantial measurement noise.



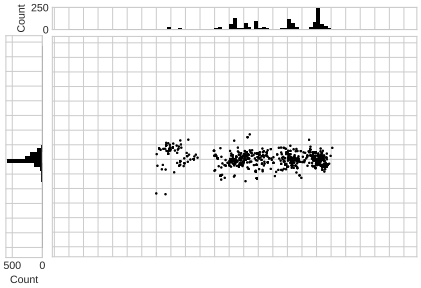
(a) Smith et al. (2004) original.



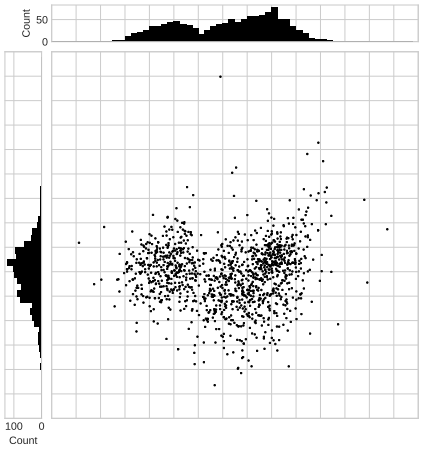
(b) Smith et al. (2004) sanitised.



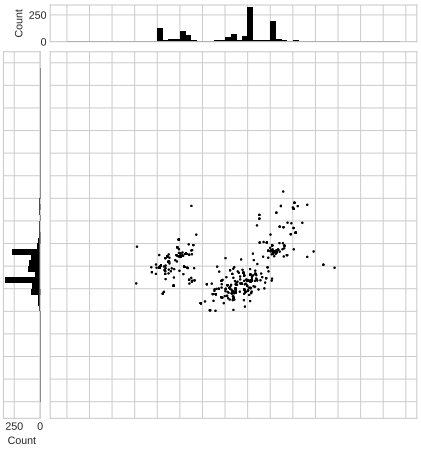
(c) Russell et al. (2008b) original.



(d) Russell et al. (2008b) sanitised.



(e) Li et al. (2016) original.



(f) Li et al. (2016) sanitised.

Figure A.3: Sanitised antigenic maps.

B

MISMATCH DISTRIBUTIONS OF SEASONS WITH MULTIPLE ANTIGENIC MAPS

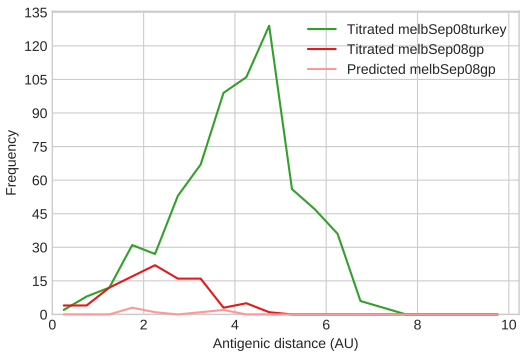


Figure B.1: Australia 2007. Turkey vs. guinea pig erythrocytes. Guinea pig erythrocytes were preferred during this period (Barr et al. 2010). Fewer titrations were conducted with guinea pig erythrocytes, but there are still a sufficient number to be confident in the distribution of mismatch distances.

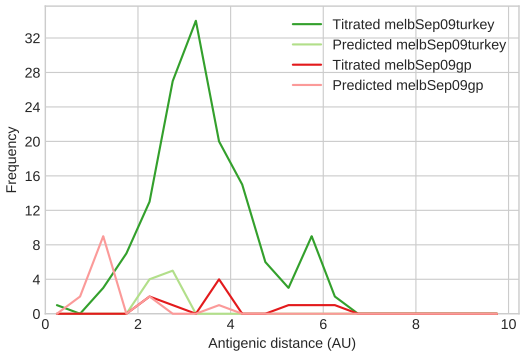


Figure B.2: Australia 2008. Turkey vs. guinea pig erythrocytes. Guinea pig erythrocytes were preferred during this period (Barr et al. 2010), however substantially fewer titrations were conducted with them. No strong evidence suggests these distributions are different. The turkey map was used due to better sampling.

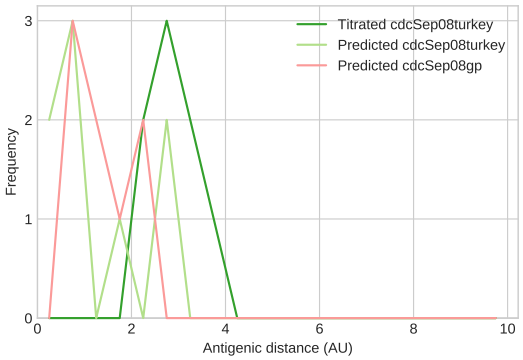


Figure B.3: Canada 2005–2006. Turkey vs. guinea pig erythrocytes. The sampling is too low in both maps. Neither were used.

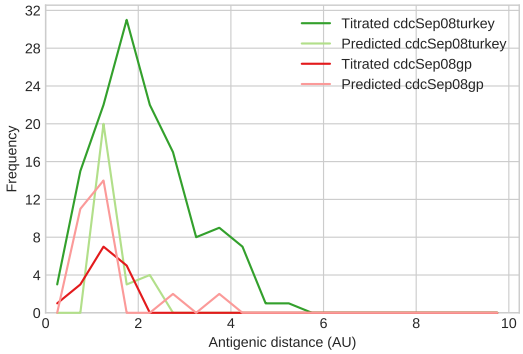


Figure B.4: Canada 2006–2007. Turkey vs. guinea pig erythrocytes. Guinea pig erythrocytes were preferred during this period (Barr et al. 2010), and were used despite lower sampling.

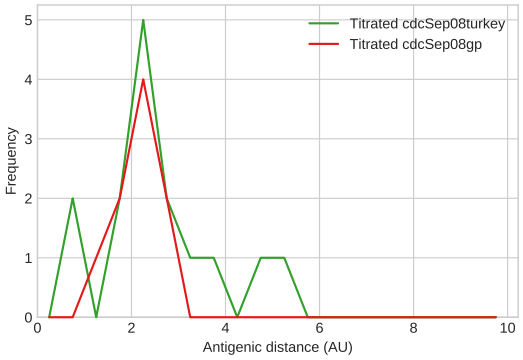


Figure B.5: Canada 2007–2008. Turkey vs. guinea pig erythrocytes. Sampling from both maps is low. Guinea pig erythrocytes were preferred during this period and were used here (Barr et al. 2010).

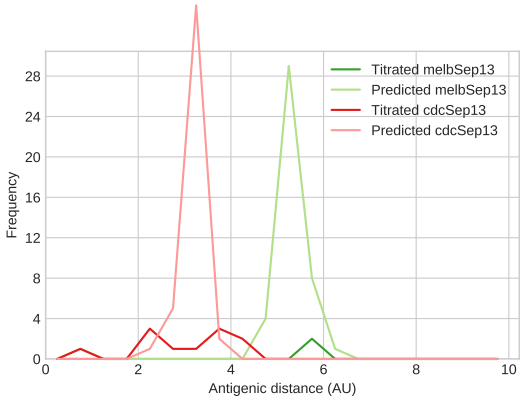


Figure B.6: China 2012–2013. VIDRL vs. CDC antigenic map. Here, the CDC map (red) has a higher number of strains titrated (11 vs. 2) and was used for that reason.

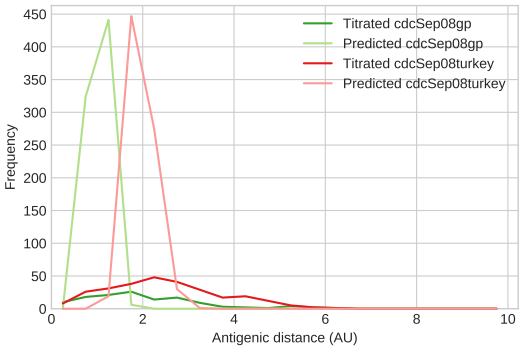


Figure B.7: U.S.A. 2007–2008. Turkey vs. guinea pig erythrocytes. Guinea pig erythrocytes were preferred during this period and were used here (Barr et al. 2010).

C

ANTIGENIC MAPS USED TO DERIVE MISMATCH MEASUREMENTS FOR ALL DATASETS

DESCRIPTION OF FIGURES Orange and green antigens circulated in the given season and country. Orange antigens were titrated, whereas green antigens were only sequenced, so their antigenic coordinates are predicted using BLUP. Blue antigens were titrated and sequenced so comprise the BLUP training data. Vaccine antigens are pink, and the mean vaccine position is shown by a cross. All other antigens in the map are shown in black. The maps also show the marginal density of each dataset to the top and right. Mismatch distributions are shown in the right panel. The lines trace the height of histogram bins which have a width of 0.5 AU.

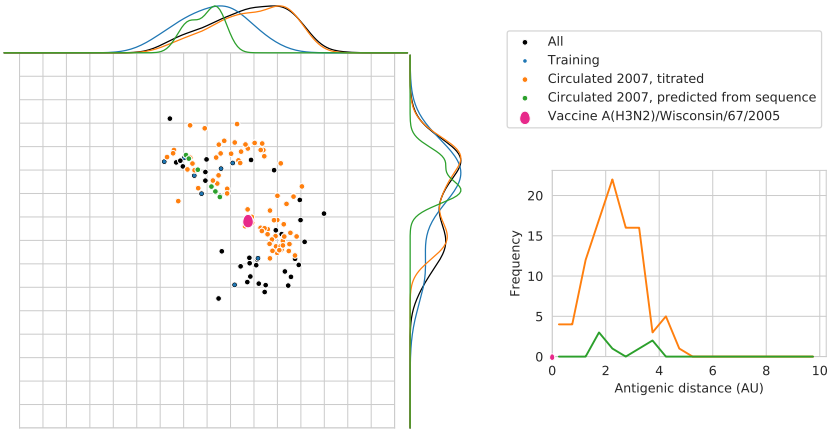


Figure C.1: Australia 2007. Dataset: melbSep08gp.

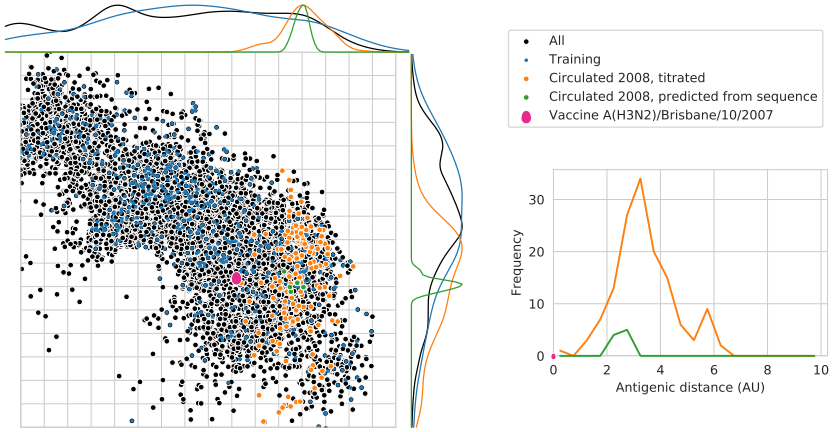


Figure C.2: Australia 2008. Dataset: melbSep09turkey.

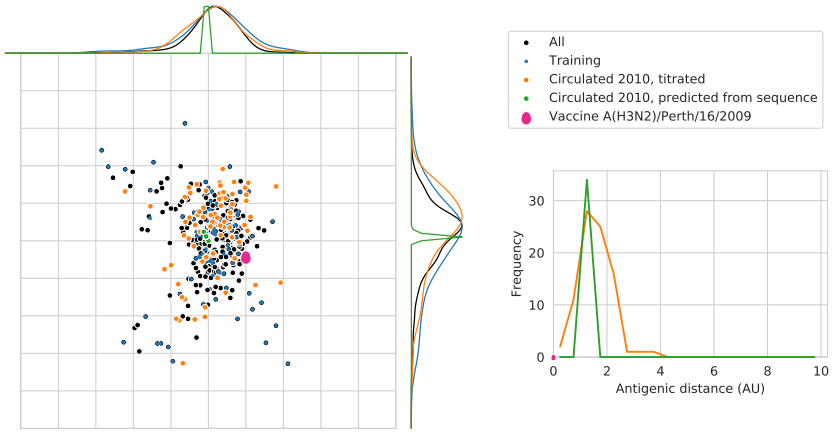


Figure C.3: Australia 2010. Dataset: melbSep11gp.

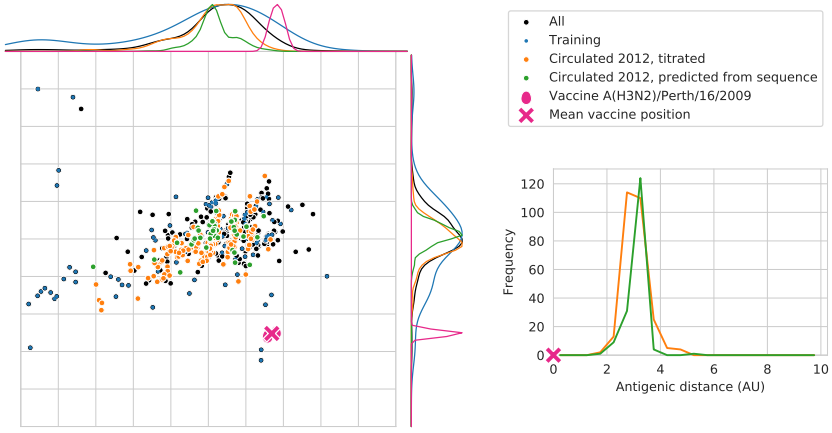


Figure C.4: Australia 2012. Dataset: melbSep13.

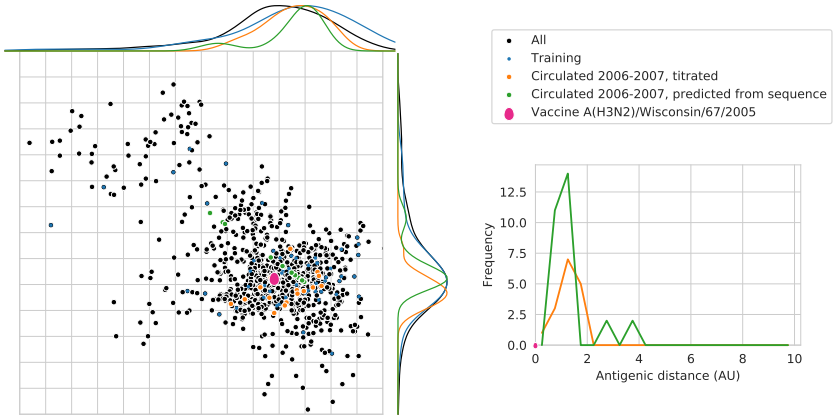


Figure C.5: Canada 2006–2007. Dataset: cdcSep08gp.

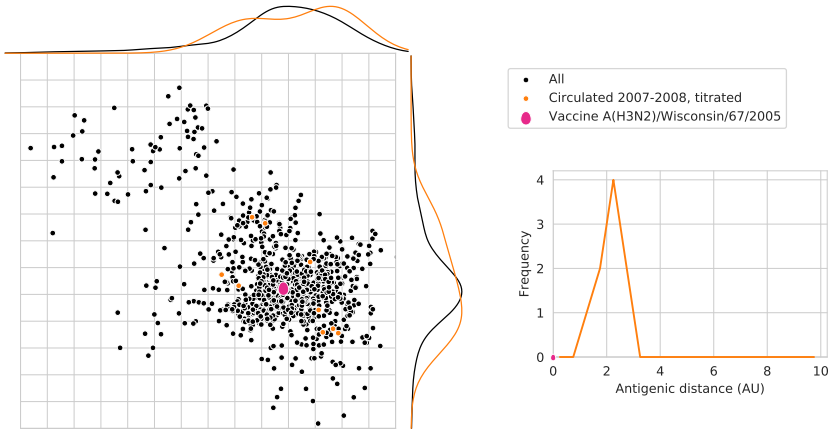


Figure C.6: Canada 2007–2008. Dataset: cdcSep08gp.

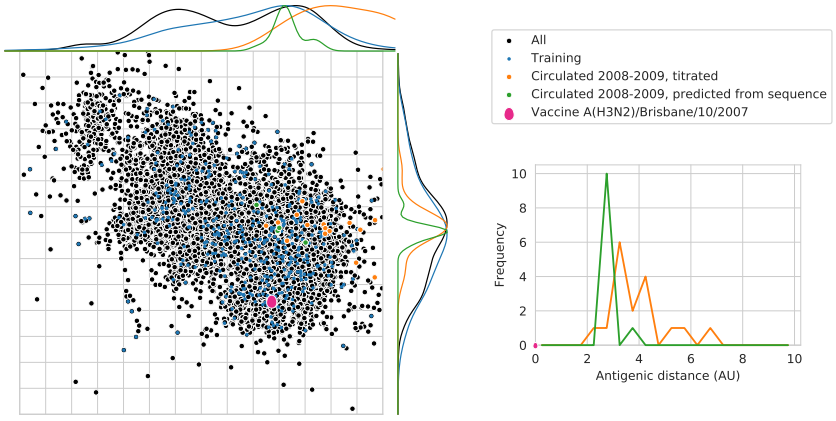


Figure C.7: Canada 2008–2009. Dataset: cdcSep09.

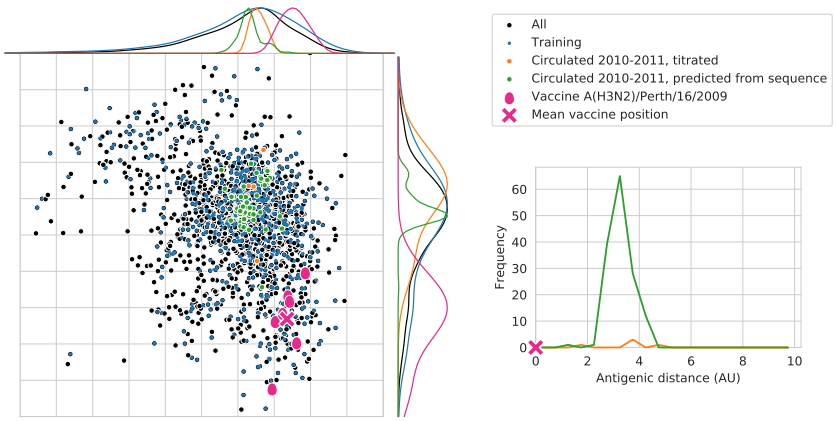


Figure C.8: Canada 2010–2011. Dataset: cdcSep11.

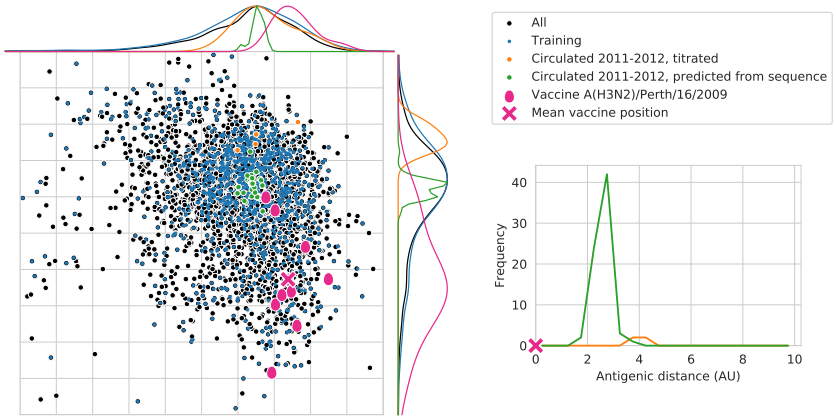


Figure C.9: Canada 2011–2012. Dataset: cdcSep12.

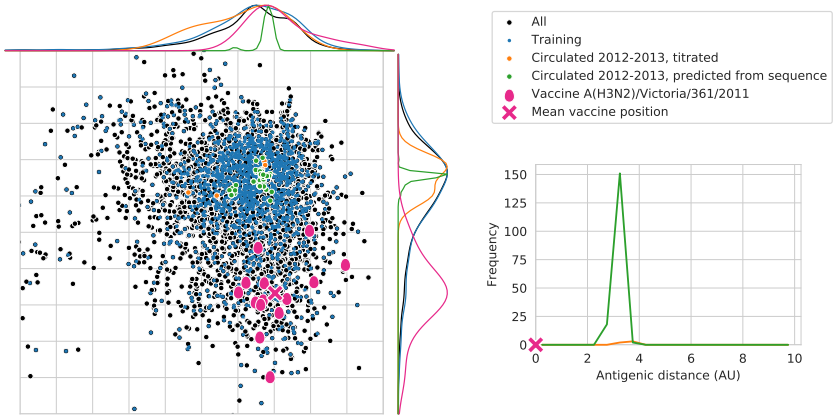


Figure C.10: Canada 2012–2013. Dataset: cdcSep13.

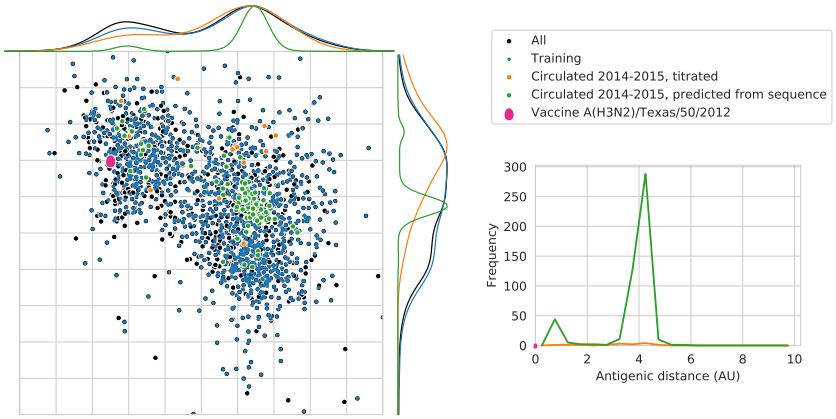


Figure C.11: Canada 2014–2015. Dataset: cdcSep15.

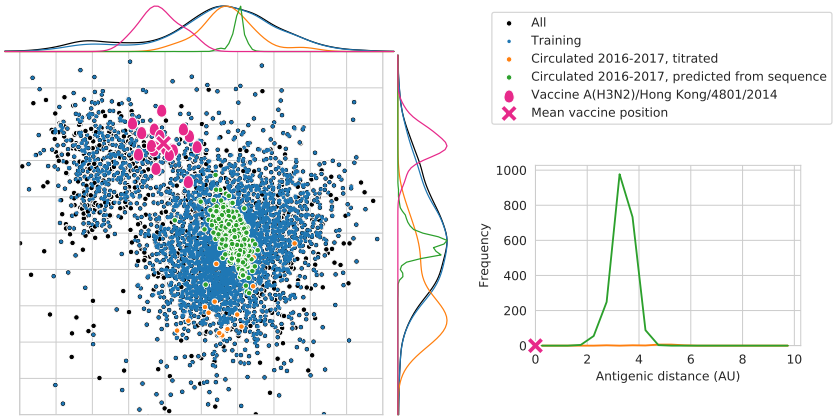


Figure C.12: Canada 2016–2017. Dataset: cdcFeb18.

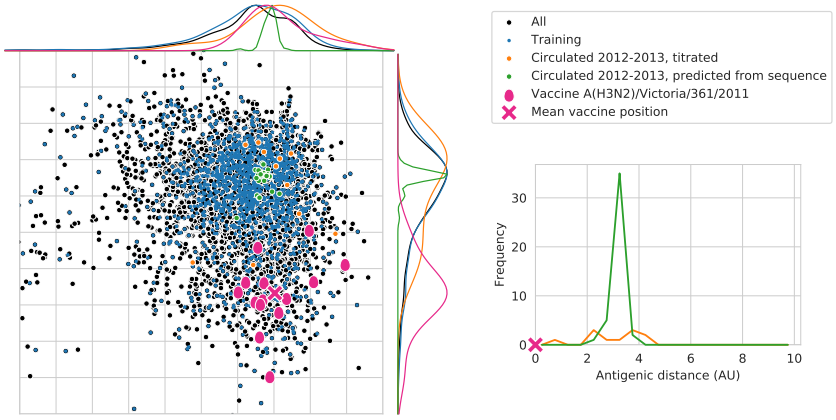


Figure C.13: China 2012–2013. Dataset: cdcSep13.

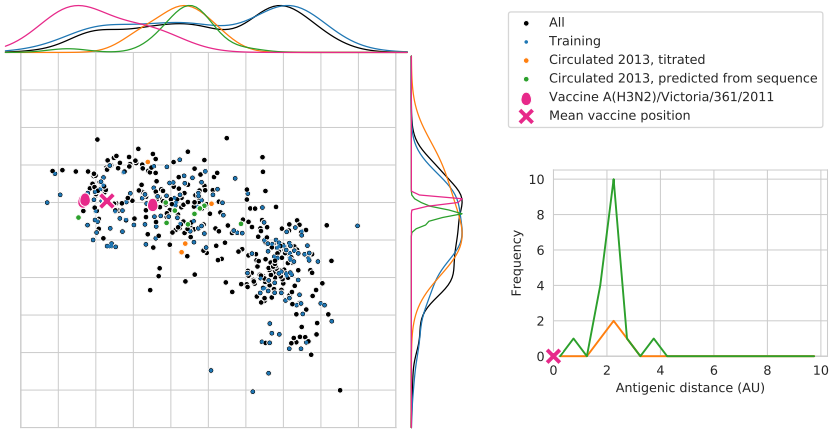


Figure C.14: New Zealand 2013. Dataset: melbSep14.

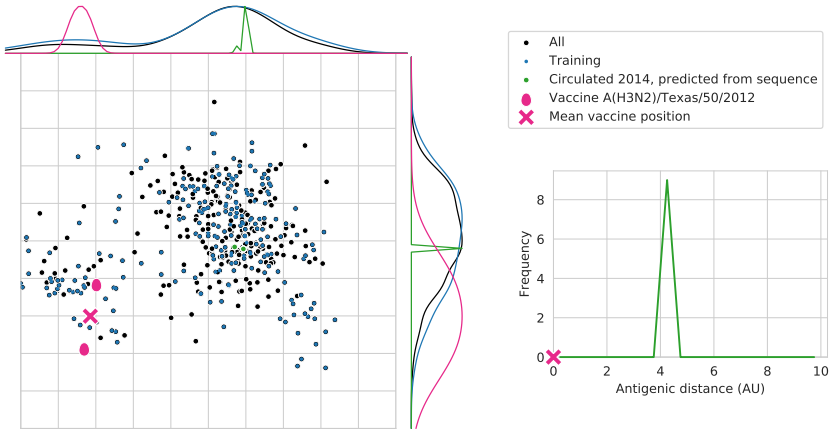


Figure C.15: South Africa 2014. Dataset: melbSep15.

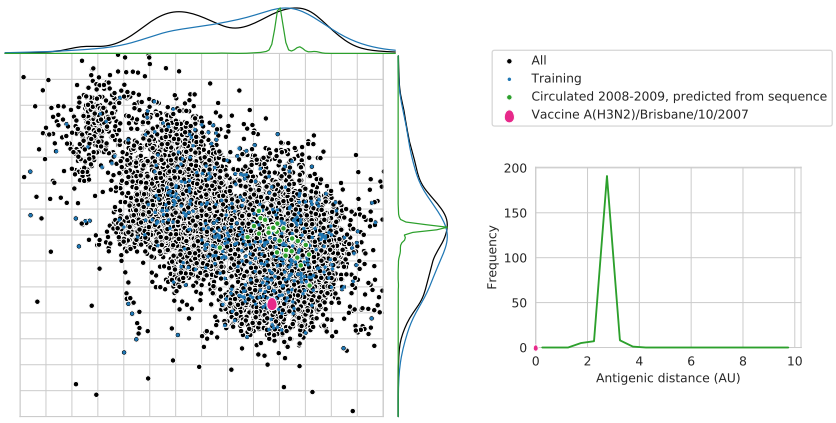


Figure C.16: Spain 2008–2009. Dataset: cdcSep09.

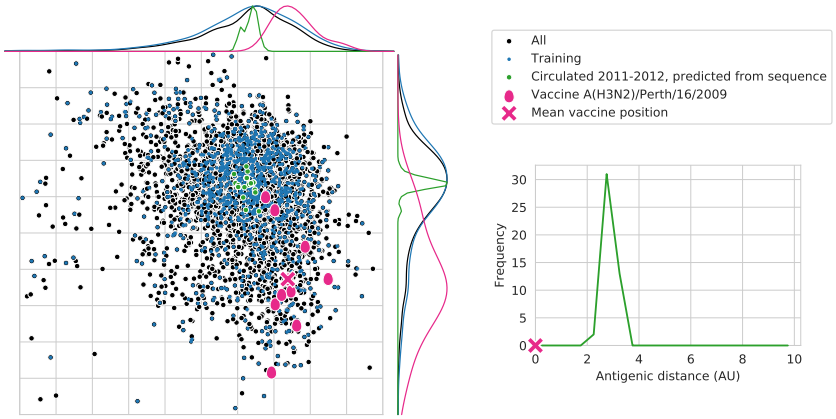


Figure C.17: Spain 2011–2012. Dataset: cdcSep12.

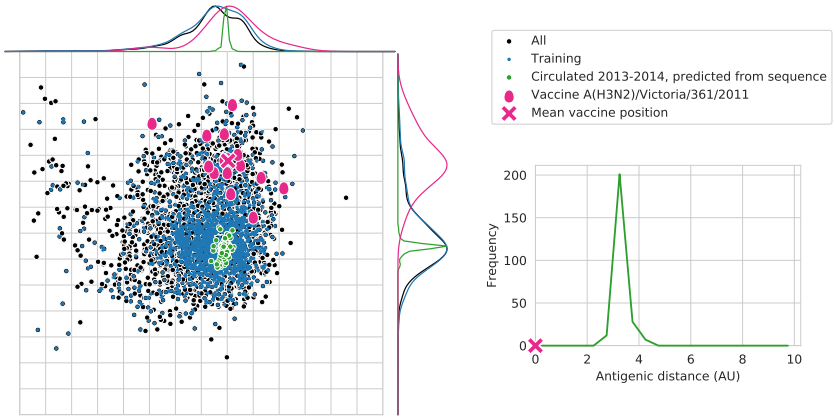


Figure C.18: Spain 2013–2014. Dataset: cdcSep14.

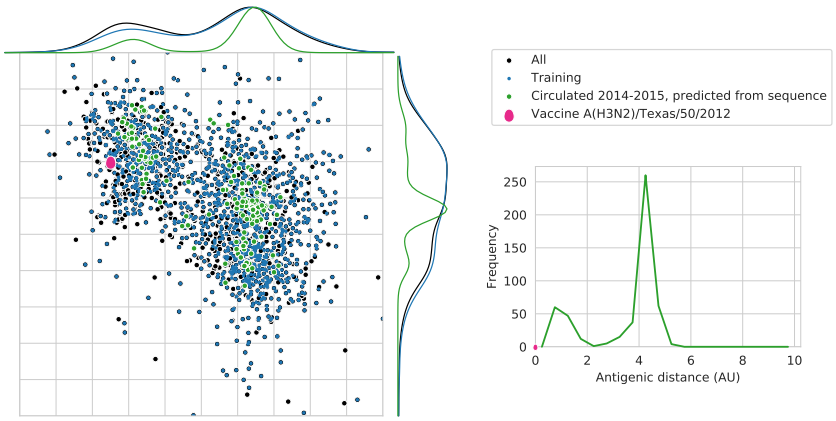


Figure C.19: Spain 2014–2015. Dataset: cdcSep15.

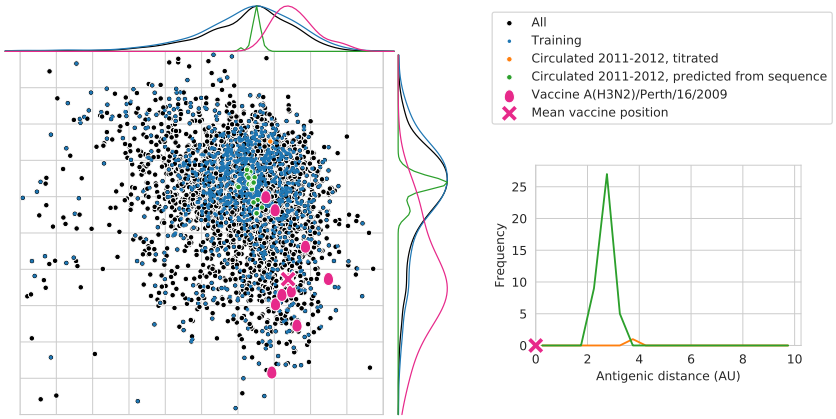


Figure C.20: U.K. 2011–2012. Dataset: cdcSep12.

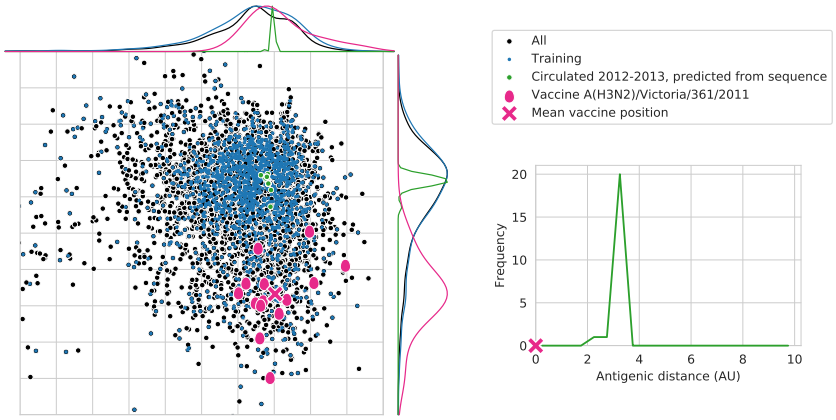


Figure C.21: U.K. 2012–2013. Dataset: cdcSep13.

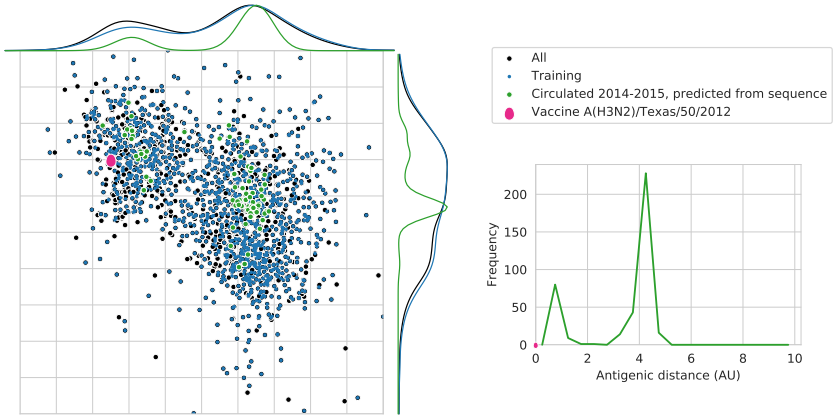


Figure C.22: U.K. 2014–2015. Dataset: cdcSep15.

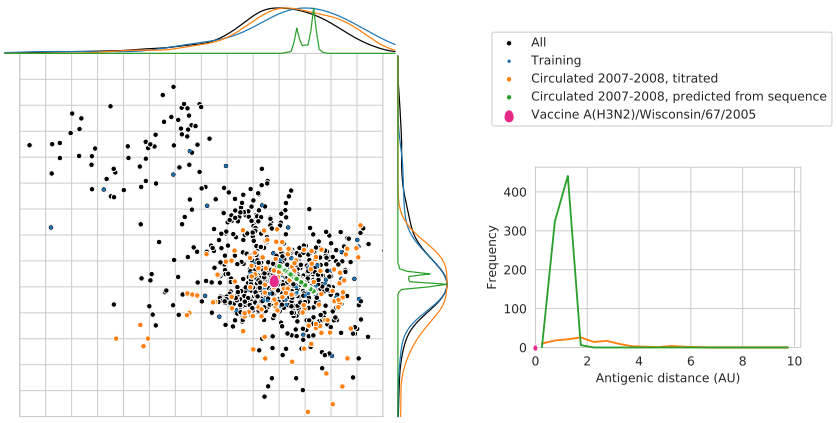


Figure C.23: U.S.A. 2007–2008. Dataset.: cdcSep08gp

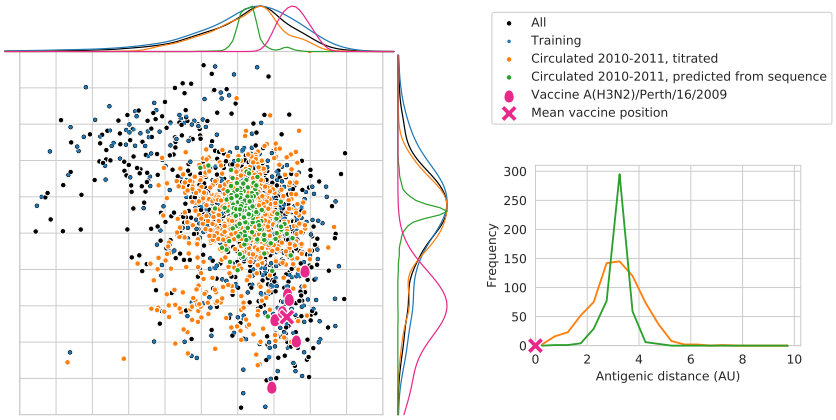


Figure C.24: U.S.A. 2010–2011. Dataset: cdcSep11.

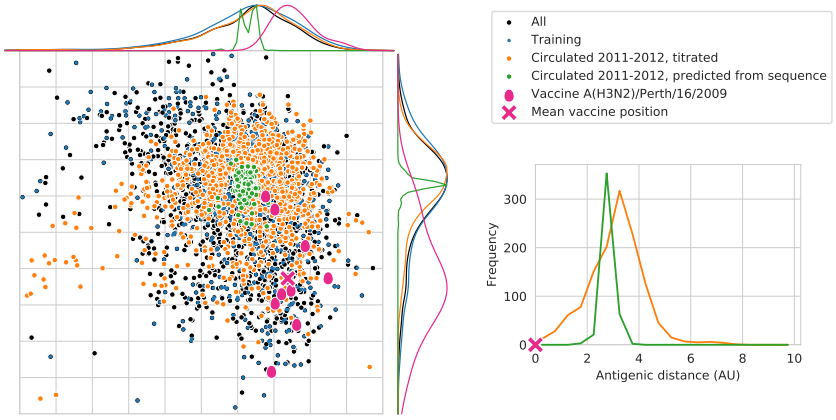


Figure C.25: U.S.A. 2011–2012. Dataset: cdcSep12.

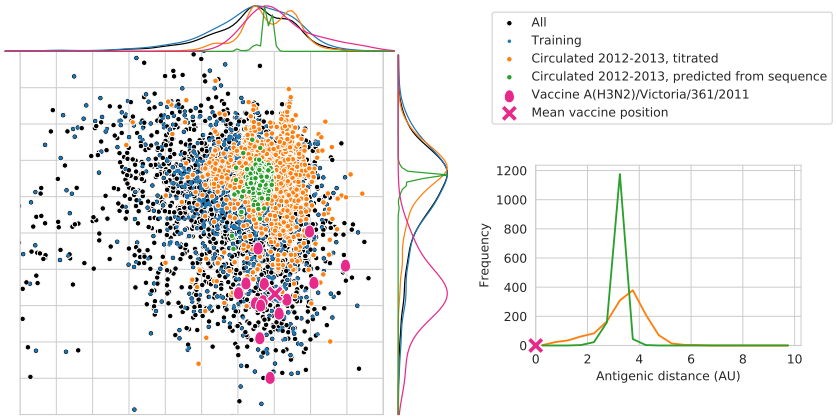


Figure C.26: U.S.A. 2012–2013. Dataset: cdcSep13.

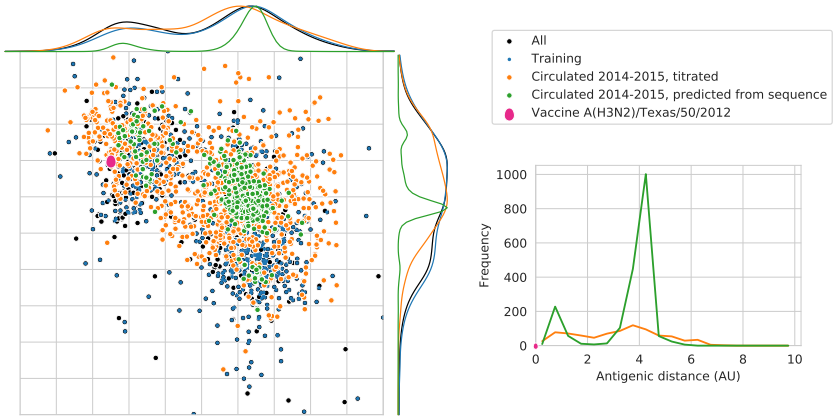


Figure C.27: U.S.A. 2014–2015. Dataset: cdcSep15.

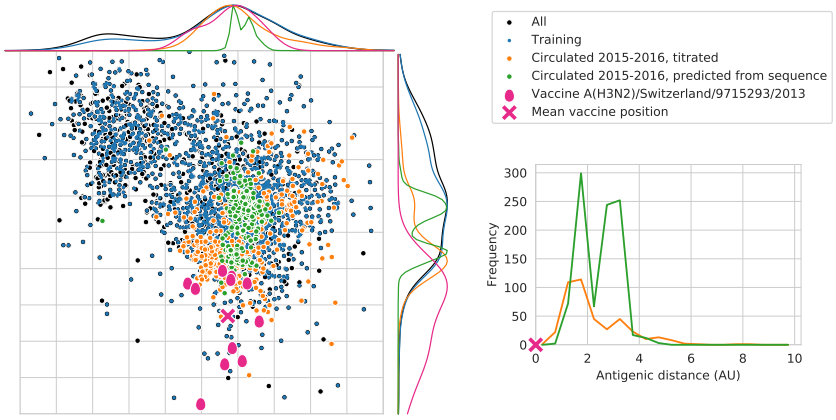


Figure C.28: U.S.A. 2015–2016. Dataset: cdcSep16.

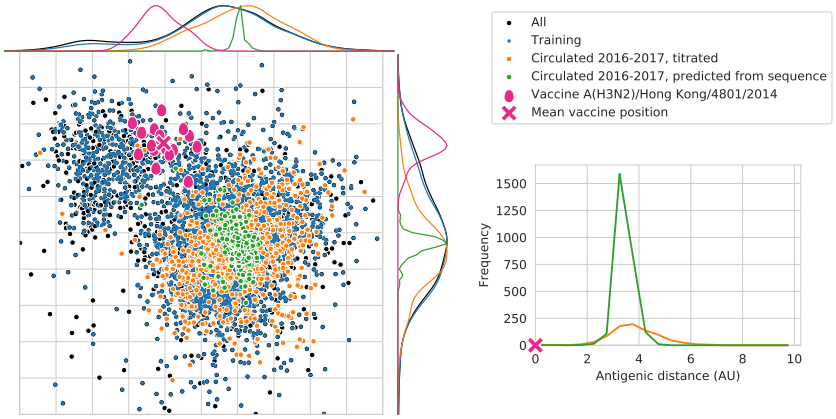


Figure C.29: U.S.A. 2016–2017. Dataset: cdcFeb18.

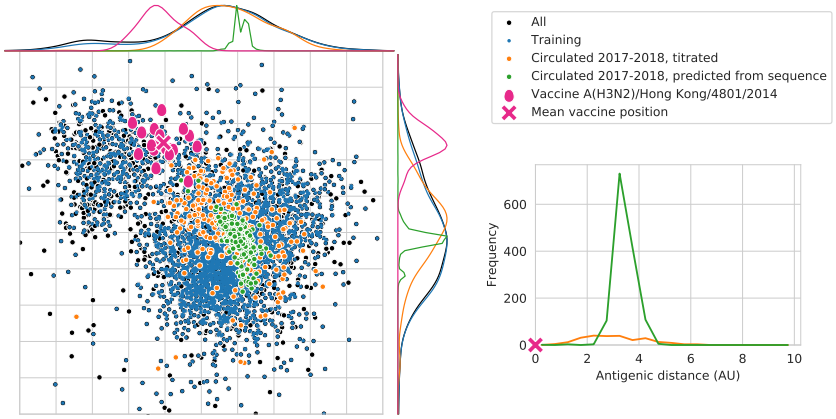


Figure C.30: U.S.A. 2017–2018. Dataset: cdcFeb18.

D

KERNEL DENSITY ESTIMATES USED TO DERIVE MISMATCH DISTRIBUTIONS FOR ALL DATASETS

DESCRIPTION OF FIGURES The top panels show strains that circulated in a country in a given influenza season. The bottom panels show vaccine antigens used in this country and season. Antigenic coordinates of individual strains are shown with grey dots, contours display the probability density function of the KDE of each dataset. Left panels show a top-down view of the KDE. Right panels show a 3D projection with the shape of the KDE also projected onto the margins.

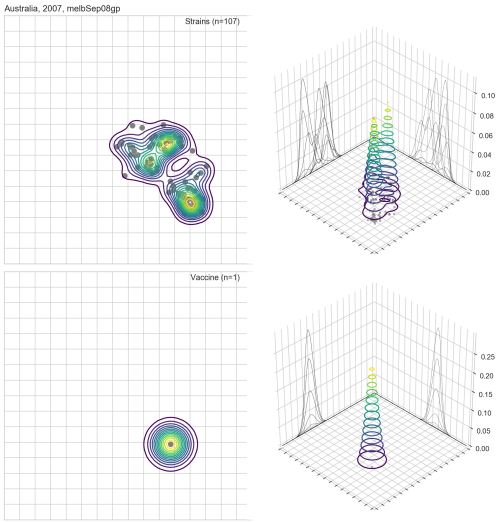


Figure D.1: Australia 2007.
Dataset: melbSep08gp.

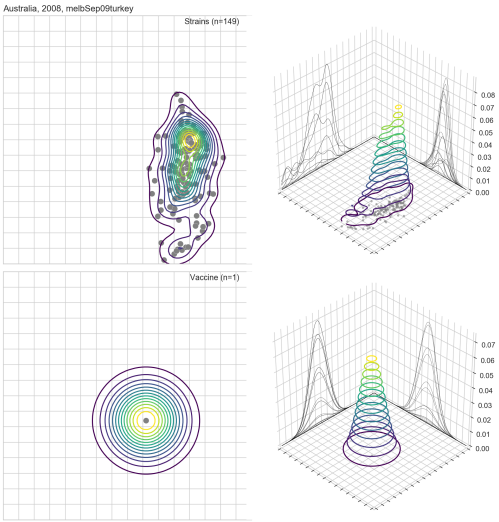


Figure D.2: Australia 2008.
Dataset: melbSep09turkey.

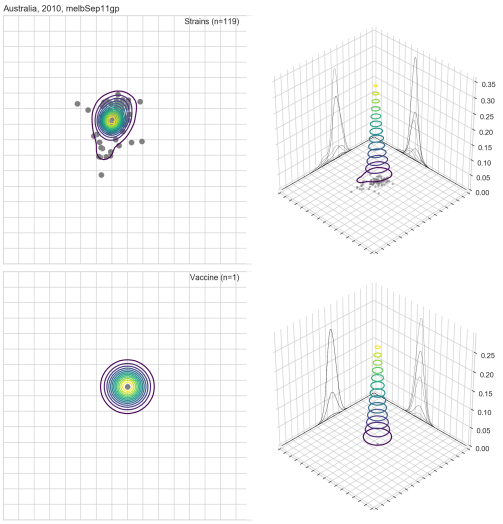


Figure D.3: Australia 2010.
Dataset: melbSep11gp.

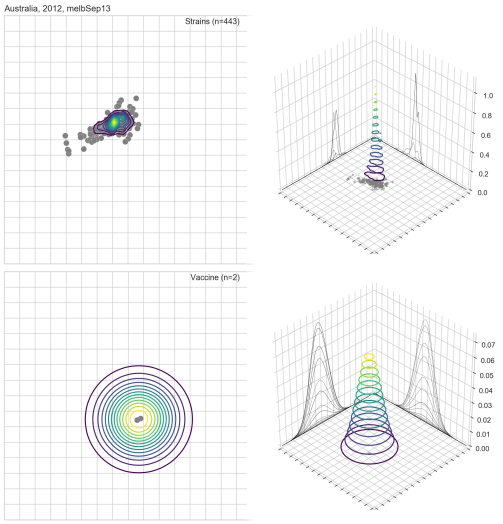


Figure D.4: Australia 2012.
Dataset: melbSep13.

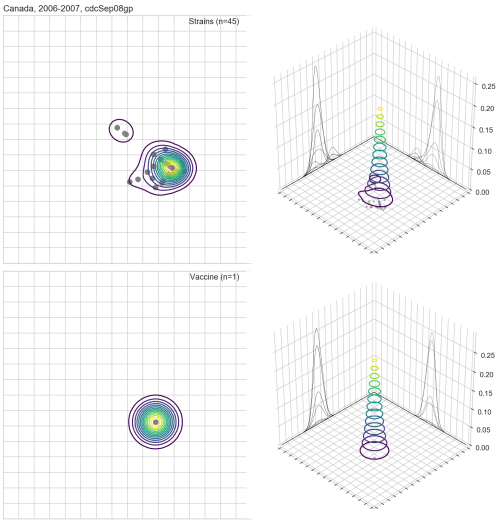


Figure D.5: Canada 2006–2007.
Dataset: cdcSep08gp.

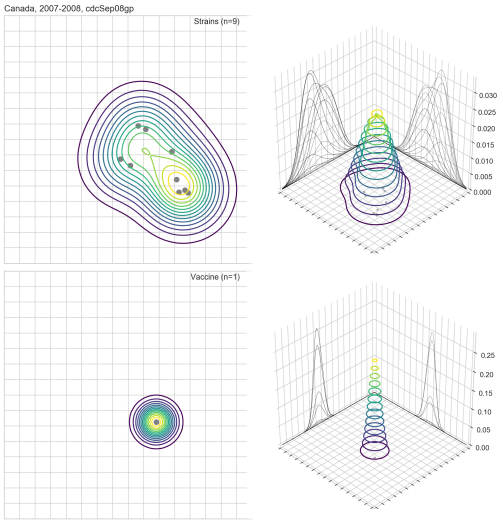


Figure D.6: Canada 2007–2008.
Dataset: cdcSep08gp.

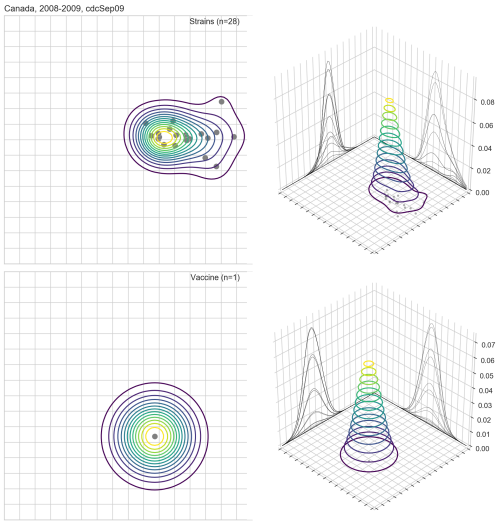


Figure D.7: Canada 2008–2009. Dataset: cdcSep09.

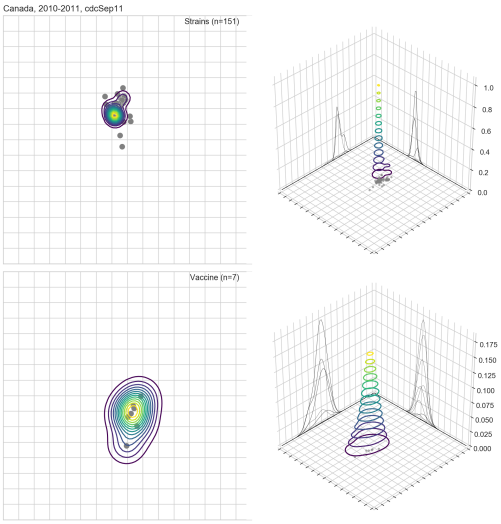


Figure D.8: Canada 2010–2011. Dataset: cdcSep11.

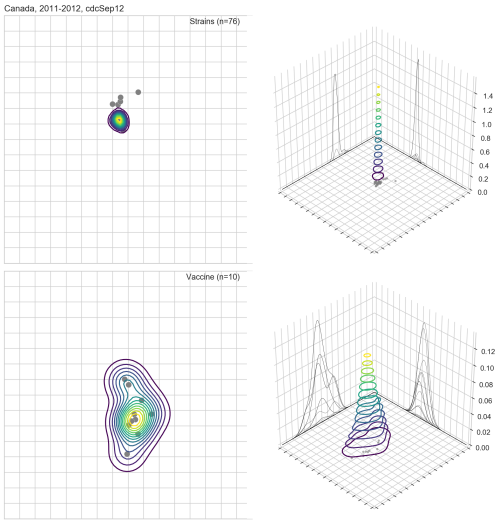


Figure D.9: Canada 2011–2012. Dataset: cdcSep12.

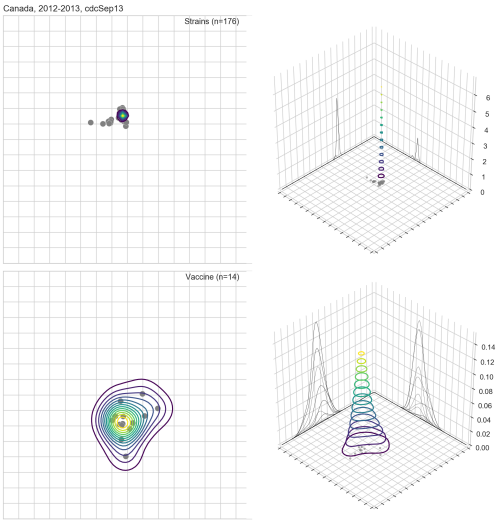


Figure D.10: Canada 2012–2013. Dataset: cdcSep13.

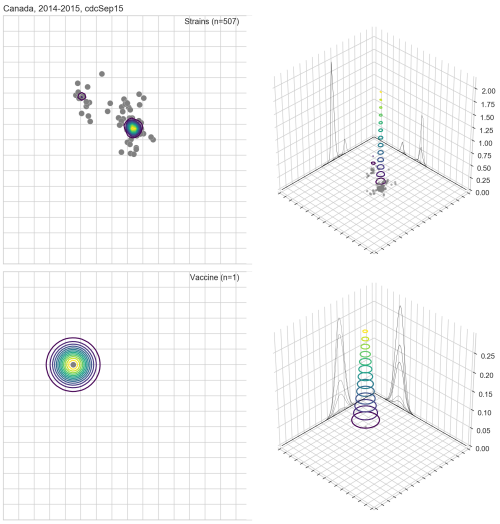


Figure D.11: Canada 2014–2015. Dataset: cdcSep15.

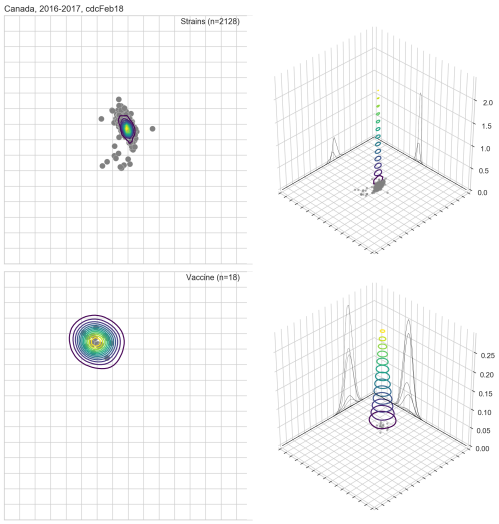


Figure D.12: Canada 2016–2017. Dataset: cdcFeb18.

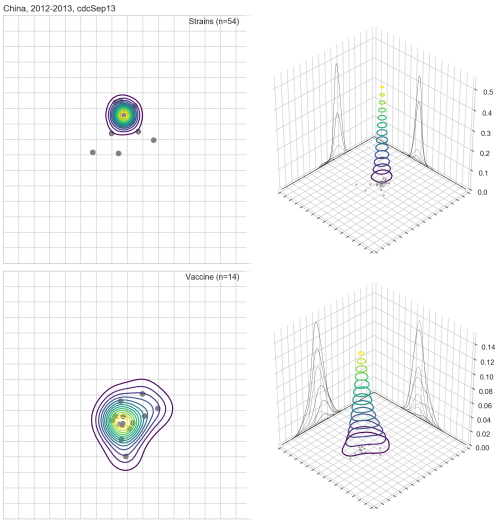


Figure D.13: China 2012–2013. Dataset: cdcSep13.

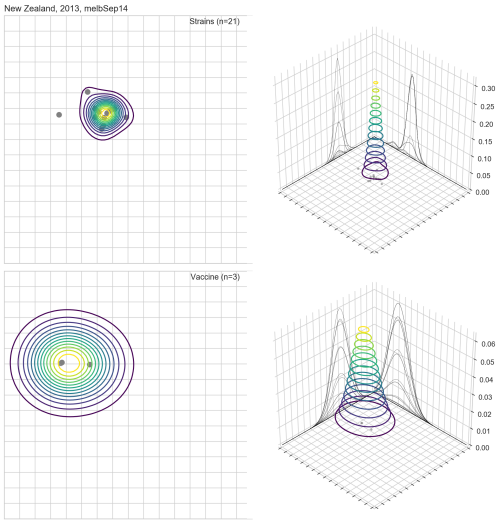


Figure D.14: New Zealand 2013. Dataset: melbSep14.

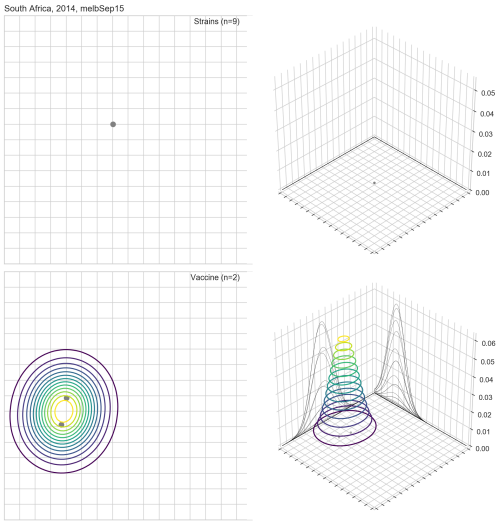


Figure D.15: South Africa 2014.
Dataset: melbSep15.

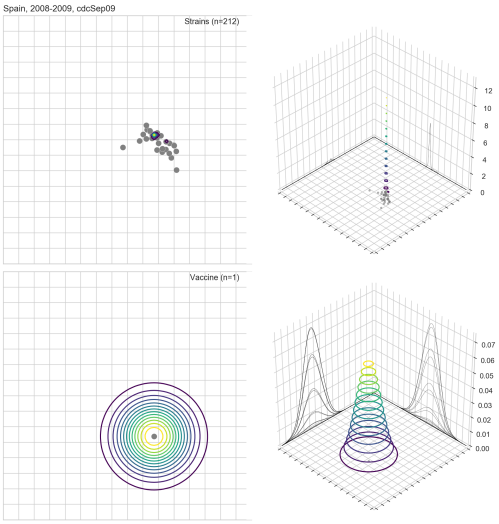


Figure D.16: Spain 2008–2009.
Dataset: cdcSep09.

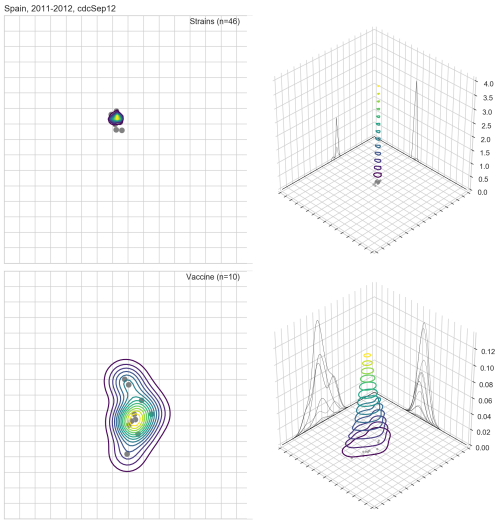


Figure D.17: Spain 2011–2012.
Dataset: cdcSep12.

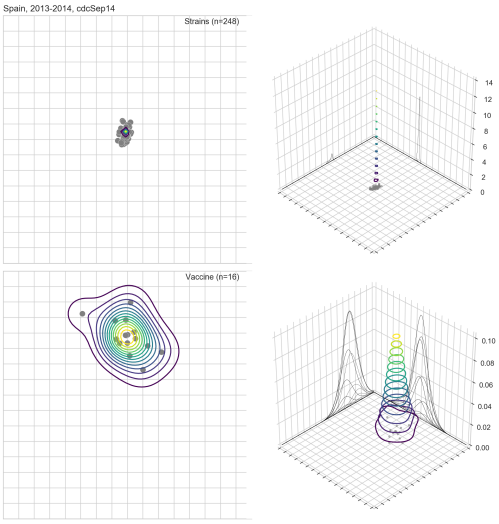


Figure D.18: Spain 2013–2014.
Dataset: cdcSep14.

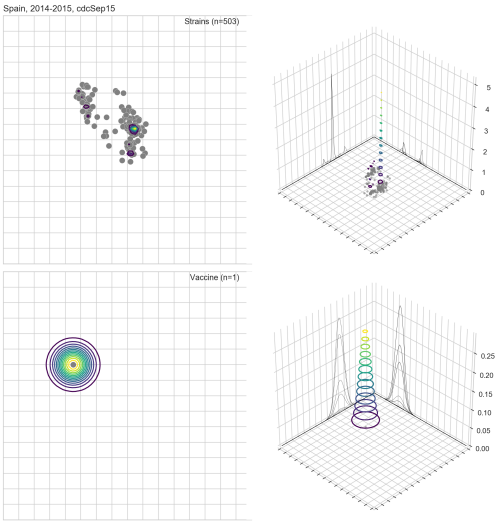


Figure D.19: Spain 2014–2015.
Dataset: cdcSep15.

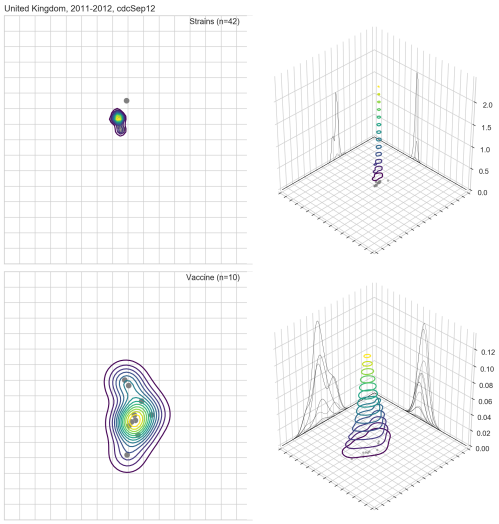


Figure D.20: U.K. 2011–2012.
Dataset: cdcSep12.

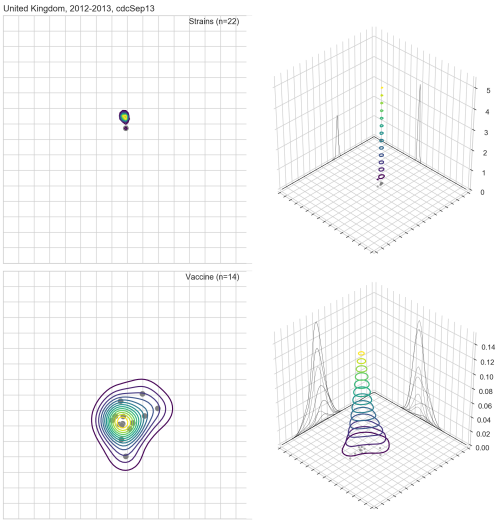


Figure D.21: U.K. 2012–2013.
Dataset: cdcSep13.

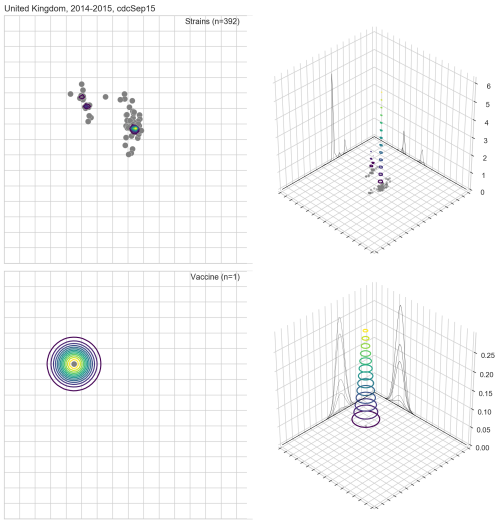


Figure D.22: U.K. 2014–2015.
Dataset: cdcSep15.

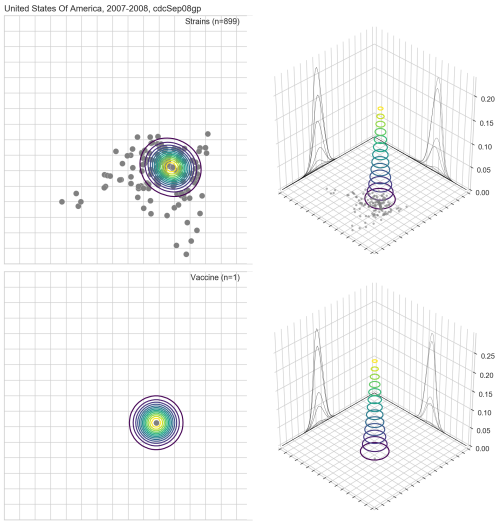


Figure D.23: U.S.A. 2007–2008. Dataset.: cdcSep08gp

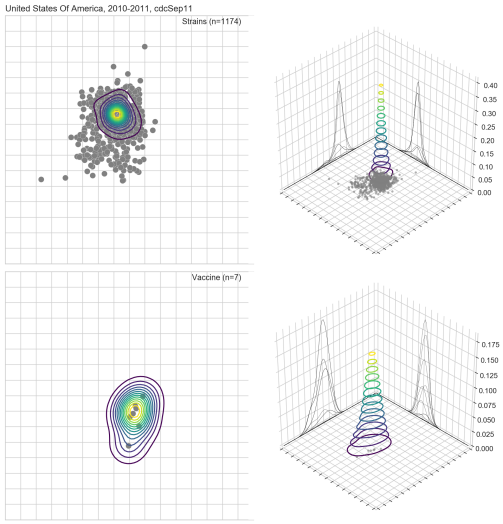


Figure D.24: U.S.A. 2010–2011. Dataset: cdcSep11.

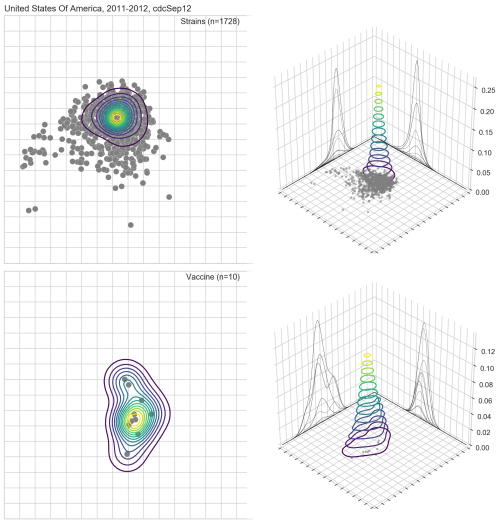


Figure D.25: U.S.A. 2011–2012. Dataset: cdcSep12.

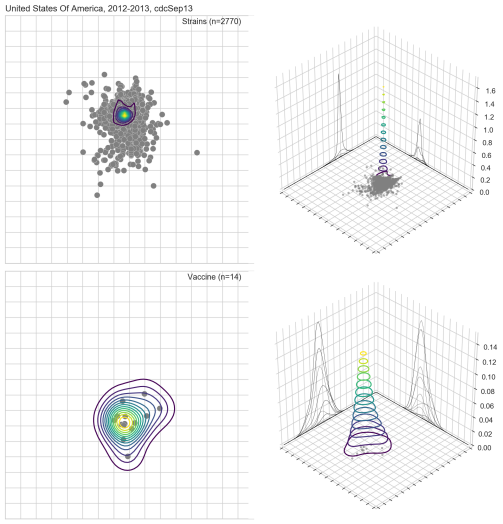


Figure D.26: U.S.A. 2012–2013. Dataset: cdcSep13.

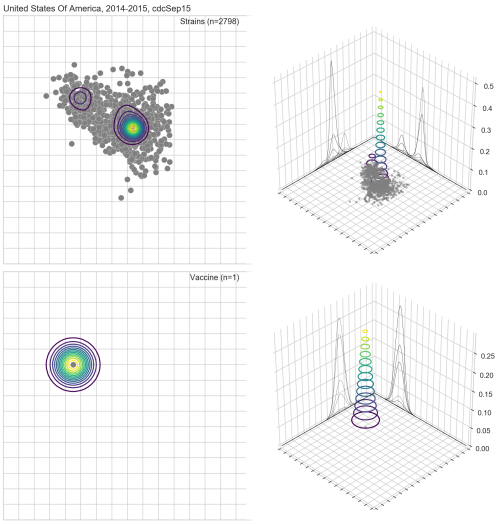


Figure D.27: U.S.A. 2014–2015. Dataset: cdcSep15.

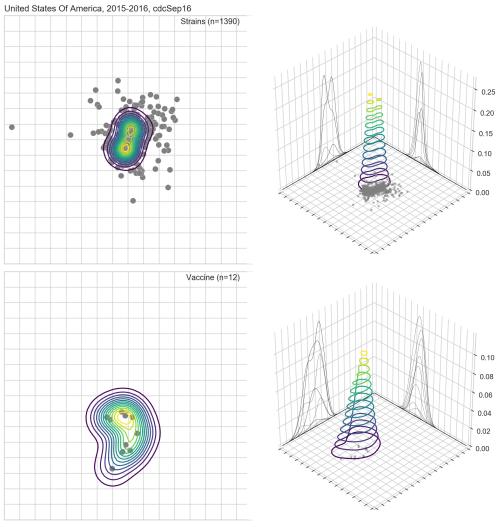


Figure D.28: U.S.A. 2015–2016. Dataset: cdcSep16.

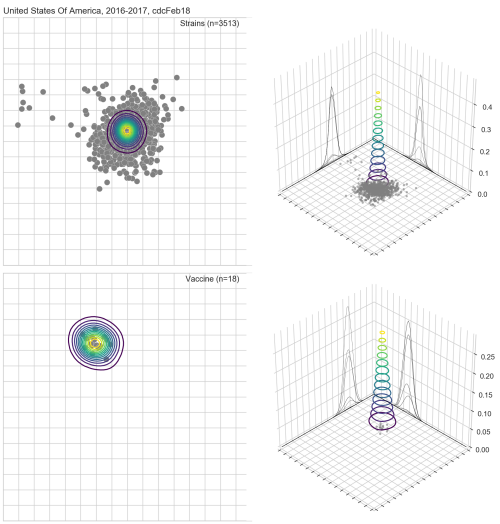


Figure D.29: U.S.A. 2016–2017. Dataset: cdcFeb18.

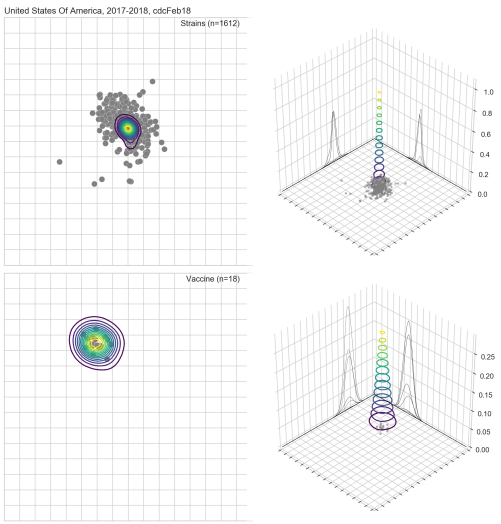


Figure D.30: U.S.A. 2017–2018. Dataset: cdcFeb18.

E

NATURAL EXPERIMENTS

E.1 FU02–CA04

E.1.1 *Strains that differ by single substitutions*

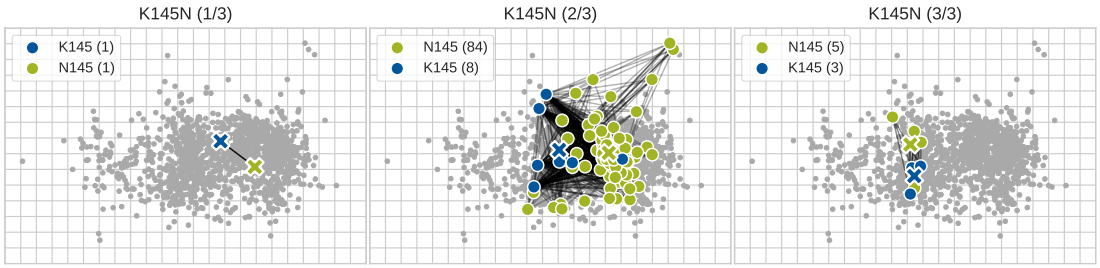


Figure E.1: K145N

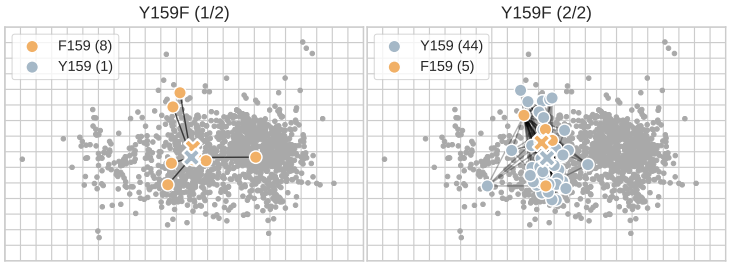


Figure E.2: Y159F

Figure E.3: V226I

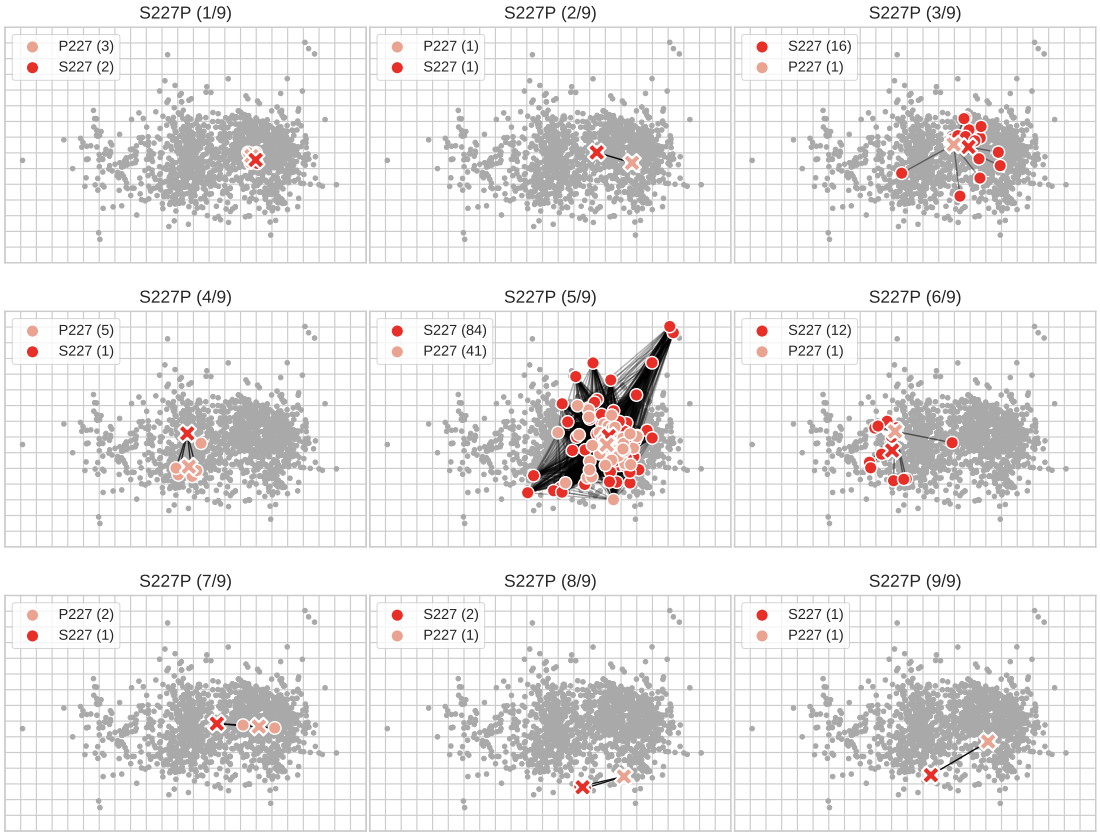


Figure E.4: S227P

E.1.2 *Individually informative strains*

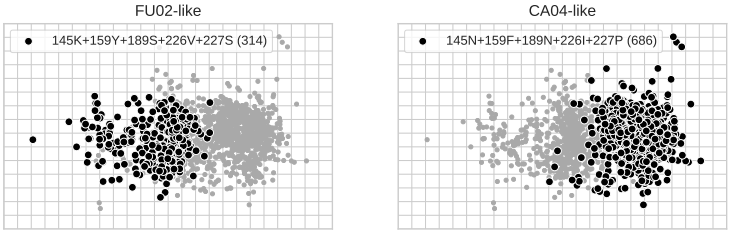


Figure E.5: Fully FU02-like and CA04-like.

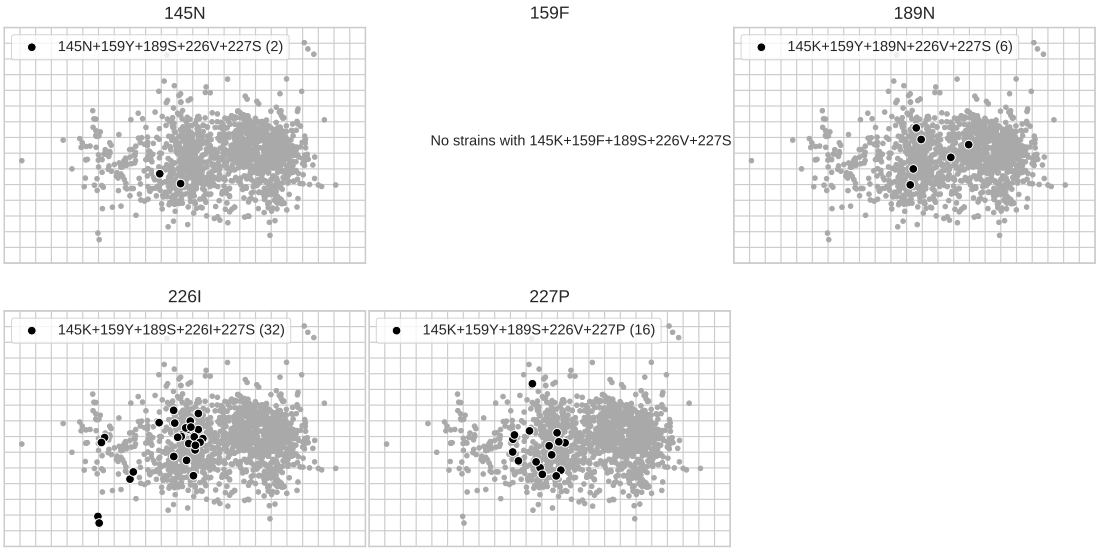


Figure E.6: FU02–CA04 single forwards substitutions.

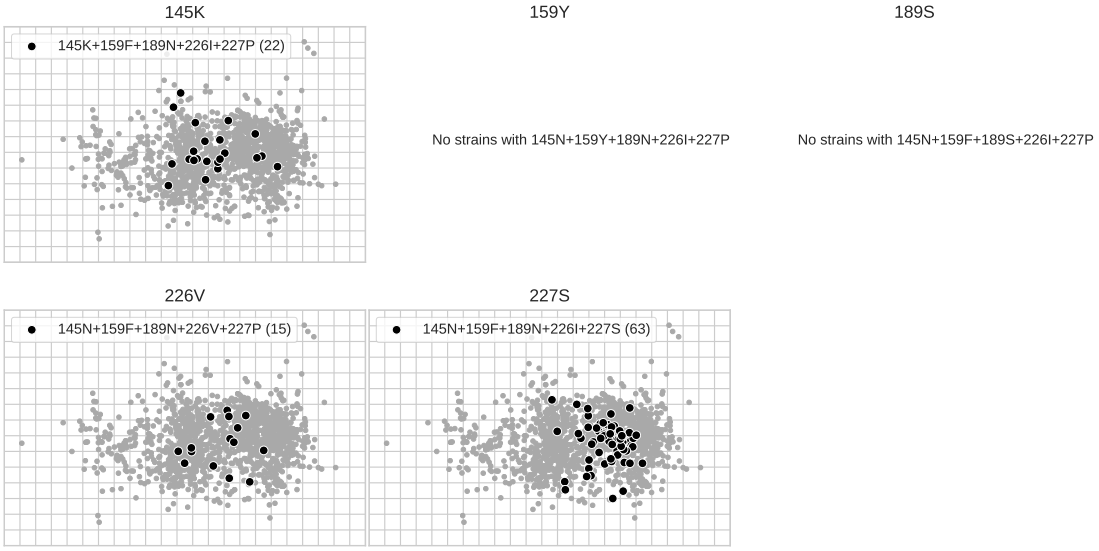


Figure E.7: FU02–CA04 single backwards substitutions.

E.2 CA04–WI05

E.2.1 *Strains that differ by single substitutions*

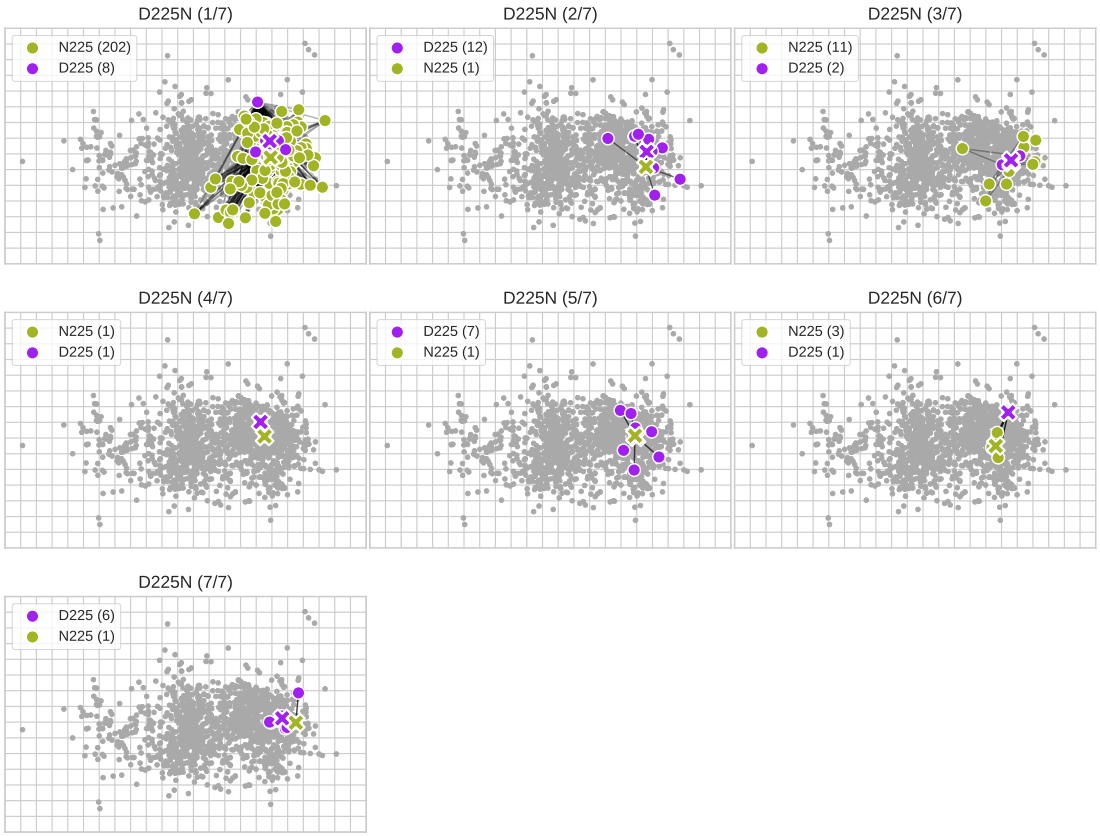


Figure E.8: D225N

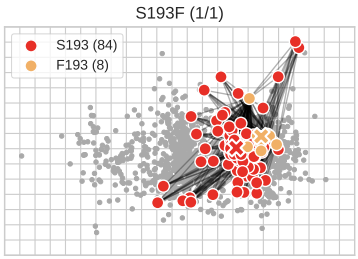


Figure E.9: S193F

E.2.2 *Individually informative strains*

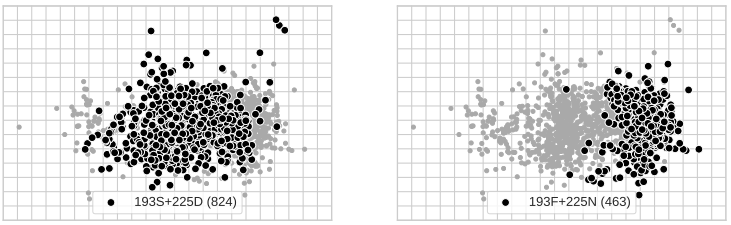


Figure E.10: Fully CA04-like and fully WI05-like.

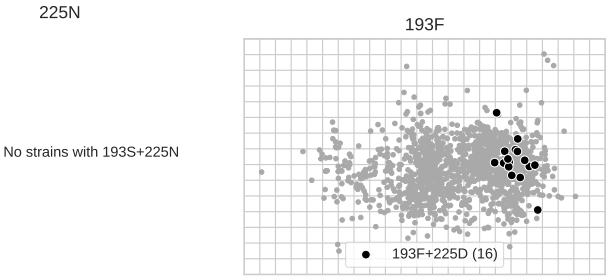


Figure E.11: CA04–WI05 single forwards substitutions.

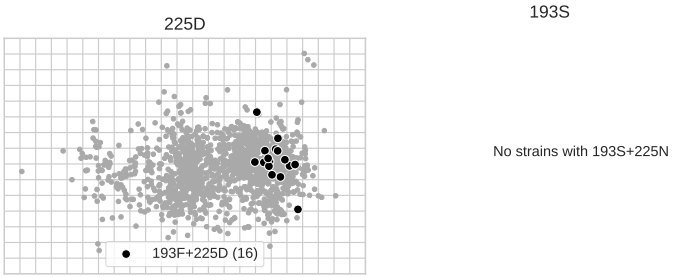


Figure E.12: CA04–WI05 single backwards substitutions.

E.3 PE09–SW13

E.3.1 *Strains that differ by single substitutions*

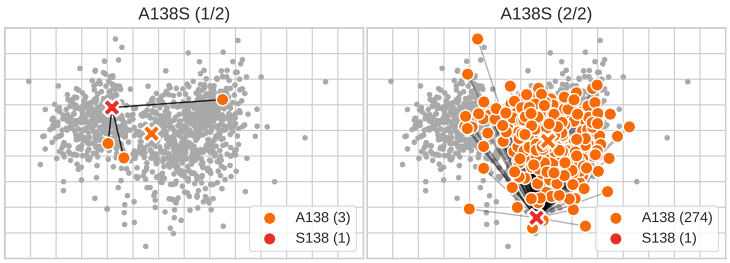


Figure E.13: A138S

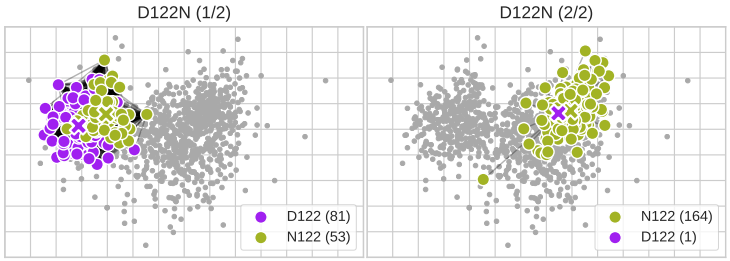


Figure E.14: D122N

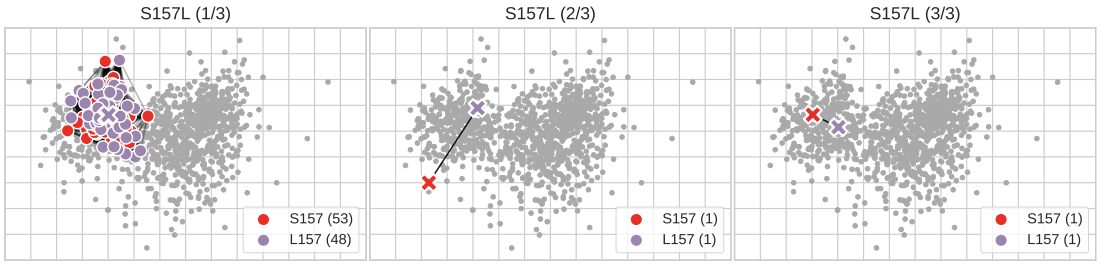


Figure E.15: S157L

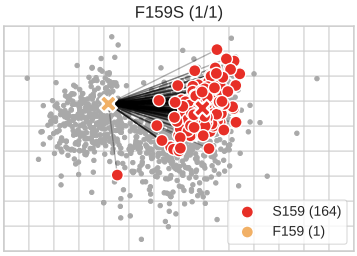


Figure E.16: F159S

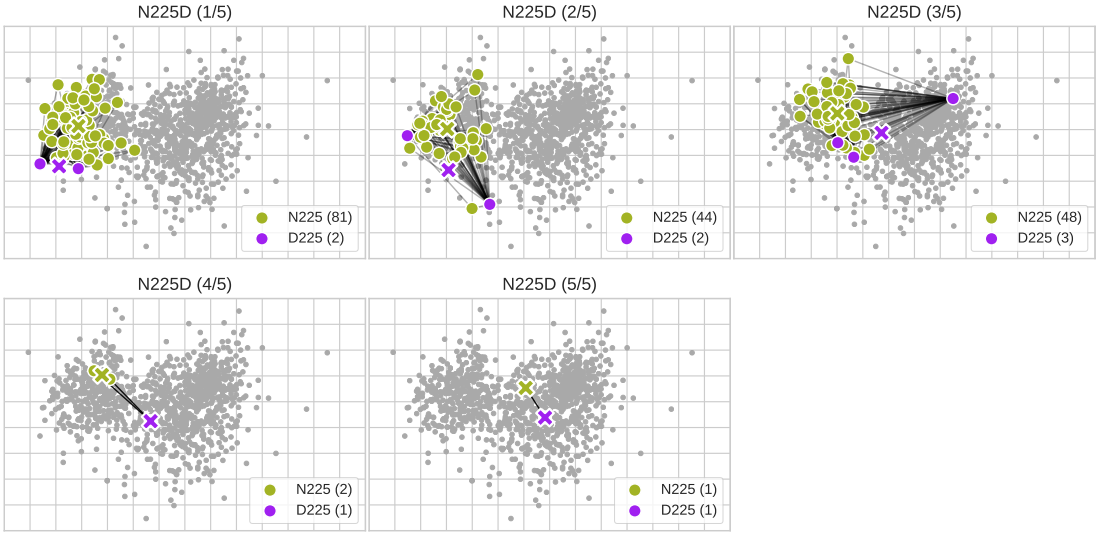


Figure E.17: N225D

E.3.2 *Individually informative strains*

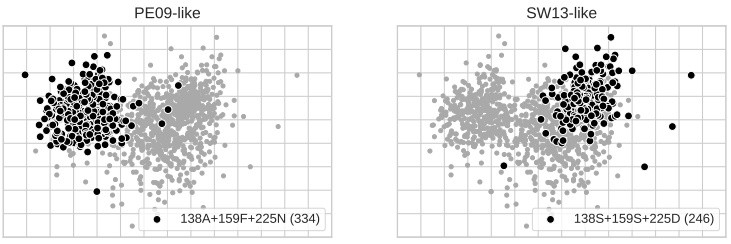


Figure E.18: Fully PE09-like and fully SW13-like.

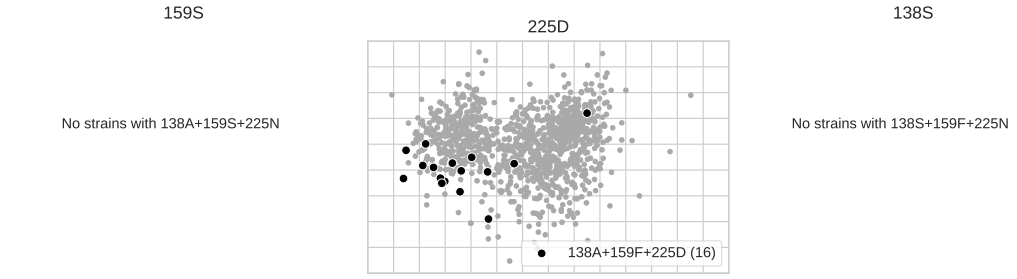


Figure E.19: PE09–SW13 single forwards substitutions.

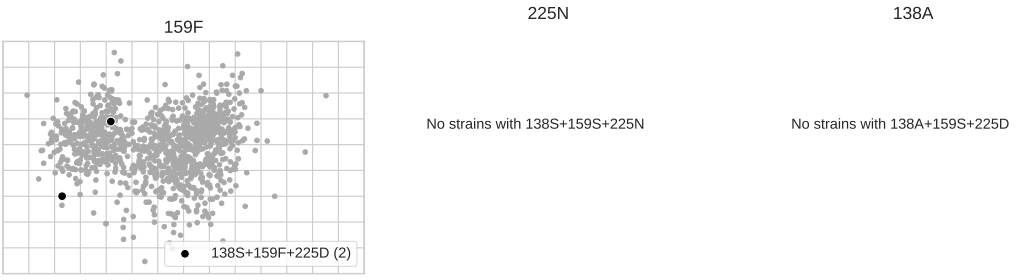


Figure E.20: PE09–SW13 single backwards substitutions.

E.4 PE09–HK14

E.4.1 *Strains that differ by single substitutions*

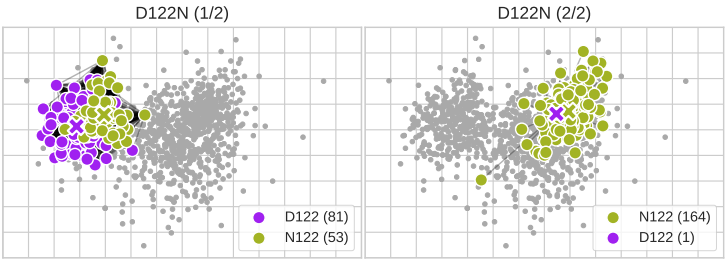


Figure E.21: D122N

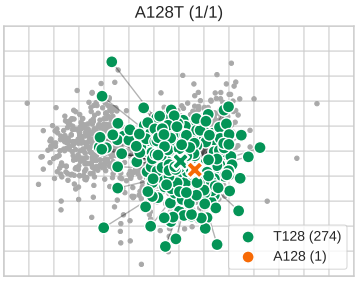


Figure E.22: A128T

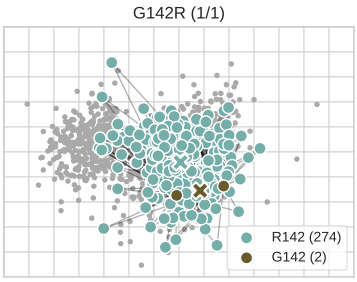


Figure E.23: G142R

E.4.2 *Individually informative strains*

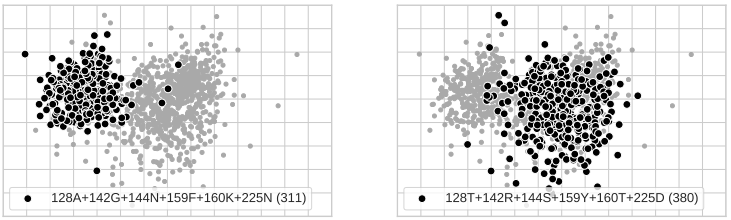


Figure E.28: Fully PE09-like and fully HK14-like.

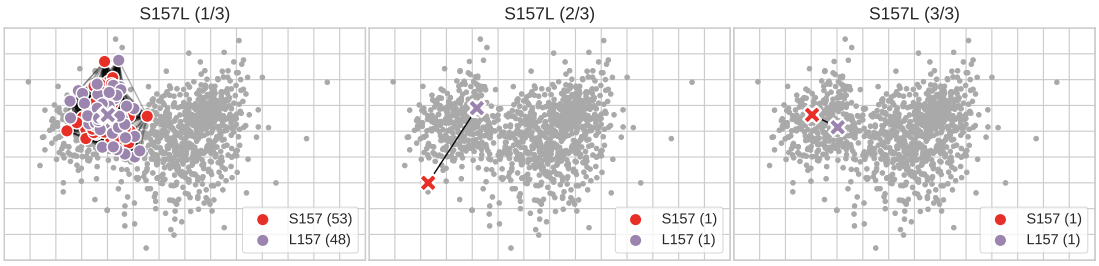


Figure E.24: S157L

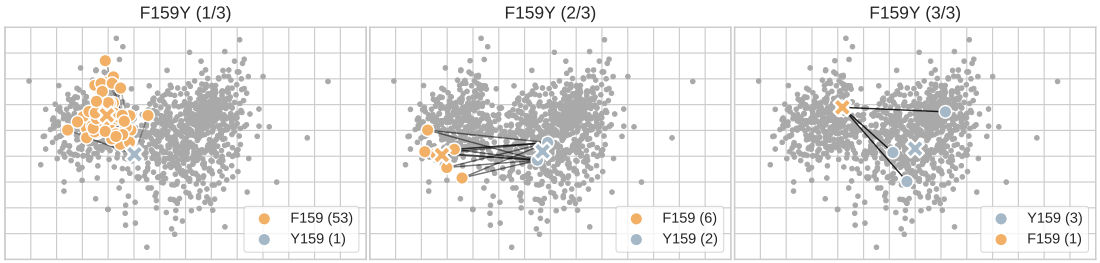


Figure E.25: F159Y

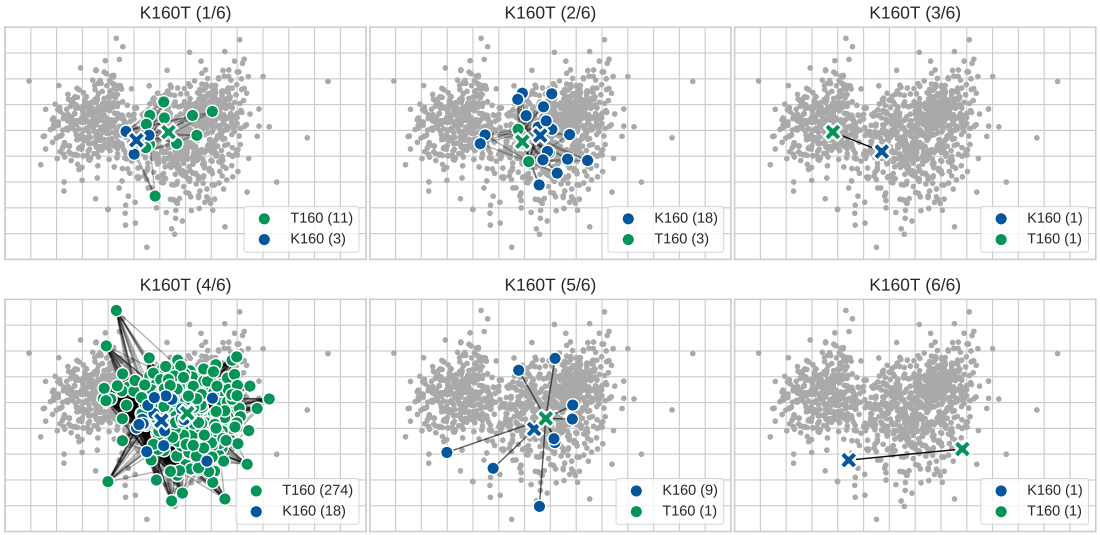


Figure E.26: K160T

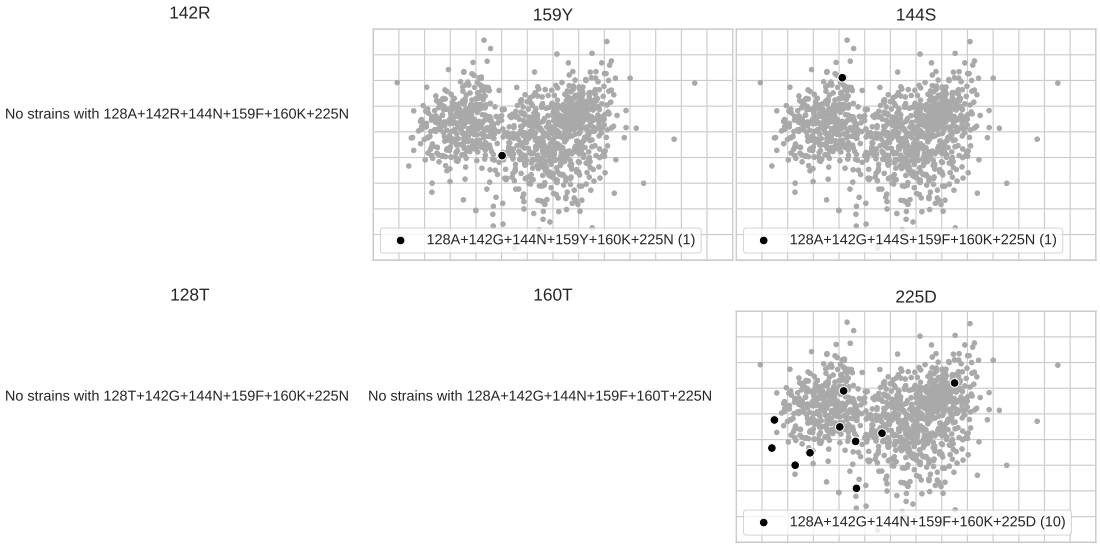


Figure E.29: PE09–HK14 single forwards substitutions.

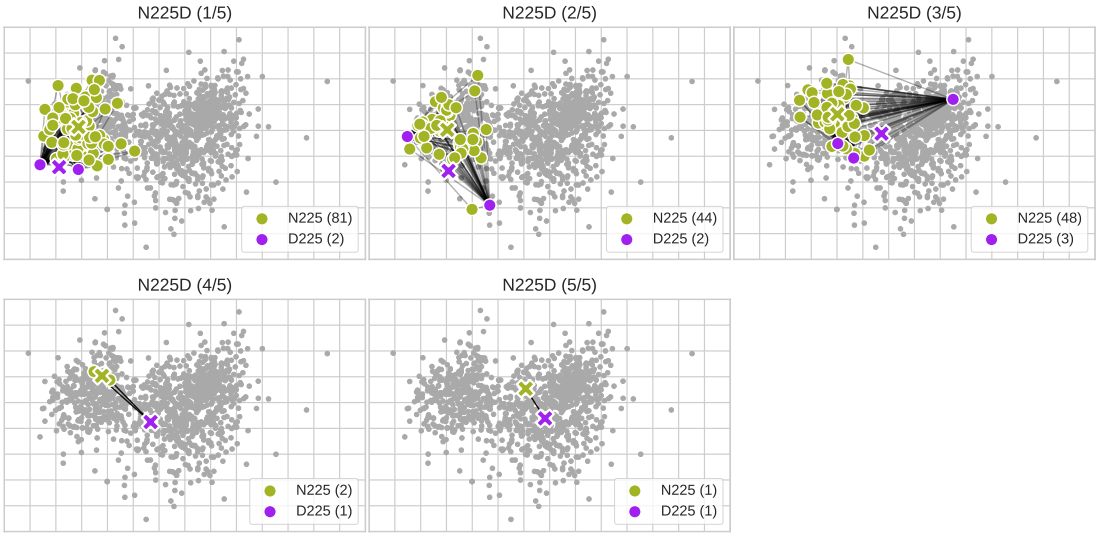


Figure E.27: N225D

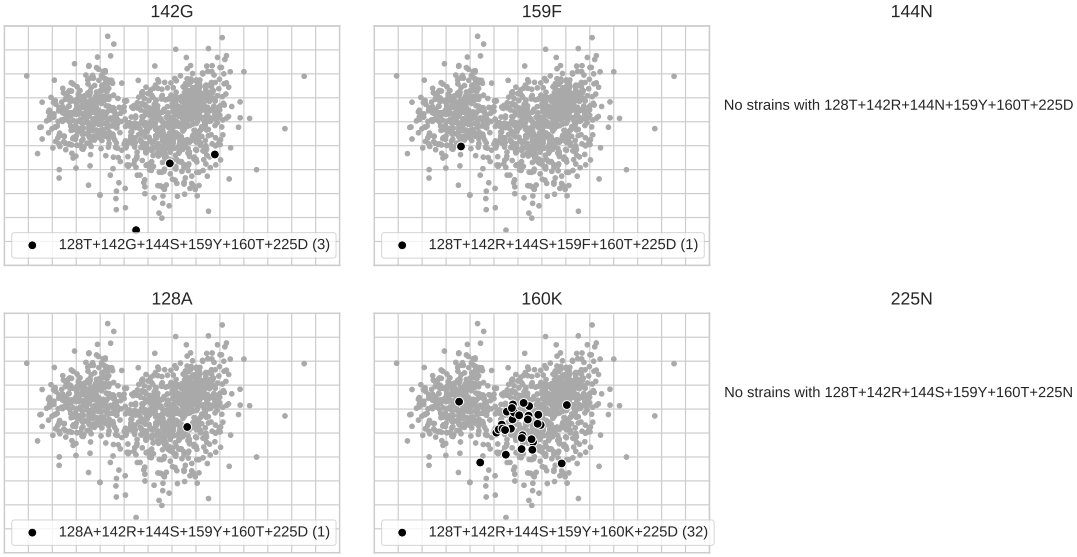


Figure E.30: PE09–HK14 single backwards substitutions.

BIBLIOGRAPHY

Abente, E. J., et al. 2016. The Molecular Determinants of Antibody Recognition and Antigenic Drift in the H3 Hemagglutinin of Swine Influenza A Virus. *Journal of Virology* 90 (18): 8266–8280. doi:10.1128/JVI.01002-16.Editor. (Cit. on pp. 13, 34, 41, 64–66, 78).

Air, G. M., and W. G. Laver. 1989. The neuraminidase of influenza virus. *Proteins: Structure, Function, and Genetics* 6 (4): 341–356. doi:10.1002/prot.340060402. (Cit. on p. 4).

Balding, D. J. 2006. A tutorial on statistical methods for population association studies. *Nature Reviews Genetics* 7 (10): 781–791. doi:10.1038/nrg1916. arXiv: 92-5-105182-8. (Cit. on p. 27).

Barman, S., et al. 2015. Egg-adaptive mutations in H3N2v vaccine virus enhance egg-based production without loss of antigenicity or immunogenicity. *Vaccine* 33 (28): 3186–3192. doi:10.1016/J.VACCINE.2015.05.011. (Cit. on p. 19).

Barr, I. G., et al. 2010. Epidemiological, antigenic and genetic characteristics of seasonal influenza A(H1N1), A(H3N2) and B influenza viruses: basis for the WHO recommendation on the composition of influenza vaccines for use in the 2009-2010 northern hemisphere season. *Vaccine* 28 (5): 1156–67. doi:10.1016/j.vaccine.2009.11.043. (Cit. on pp. 15, 123–125).

Bateson, W. 1909. *Mendel's Principles of Heredity*. Cambridge University Press. (Cit. on p. 34).

Bedford, T., et al. 2014. Integrating influenza antigenic dynamics with molecular evolution. *eLife* 3:e01914. doi:10.7554/eLife.01914. (Cit. on p. 12).

Belongia, E. A., et al. 2016. Variable influenza vaccine effectiveness by subtype: a systematic review and meta-analysis of test-negative design studies. *The Lancet Infectious Diseases* 16 (8): 942–951. doi:10.1016/S1473-3099(16)00129-8. (Cit. on pp. 49, 51, 61).

Belongia, E. A., et al. 2009. Effectiveness of Inactivated Influenza Vaccines Varied Substantially with Antigenic Match from the 2004–2005 Season to the 2006–2007 Season. *The Journal of Infectious Diseases* 199 (2): 159–167. doi:10.1086/595861. (Cit. on pp. 44–46).

Belongia, E. A., et al. 2011. Influenza vaccine effectiveness in Wisconsin during the 2007–08 season: Comparison of interim and final results. *Vaccine* 29 (38): 6558–6563. doi:10.1016/j.vaccine.2011.07.002. (Cit. on pp. 43, 51).

Belshe, R. B., et al. 2007. Live Attenuated versus Inactivated Influenza Vaccine in Infants and Young Children. *New England Journal of Medicine* 356 (7): 685–696. doi:10.1056/NEJMoa065368. (Cit. on p. 19).

Bonferroni, C. E. 1936. Teoria statistica delle classi e calcolo delle probabilità. *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze*, no. 8: 3–62. (Cit. on p. 33).

Burke, D. F., and D. J. Smith. 2014. A recommended numbering scheme for influenza A HA subtypes. *PloS one* 9 (11): e112302. doi:10.1371/journal.pone.0112302. (Cit. on pp. 49, 66, 101).

Burnet, F. M. 1940. Influenza Virus Infections of the Chick Embryo Lung. *The British Journal of Experimental Pathology* 21 (3): 147–153. (Cit. on pp. 14, 18).

Casale, F. P. 2016. "Multivariate linear mixed models for statistical genetics". Doctor of Philosophy (PhD), University of Cambridge. doi:https://doi.org/10.17863/CAM.13422. (Cit. on pp. 29, 35).

CDC. 2017. *Advisory Committee on Immunization Practices (ACIP) Summary Report*. Tech. rep. (Cit. on p. 51).

– . 2019. *Seasonal Influenza Vaccine Effectiveness, 2004-2018*. Visited on 03/15/2019. (Cit. on p. 51).

Chen, H., et al. 2018. Passage adaptation correlates with the reduced efficacy of the influenza vaccine. *Clinical Infectious Diseases*. doi:10.1093/cid/ciy1065. (Cit. on p. 60).

Chen, Z., H. Zhou, and H. Jin. 2010. The impact of key amino acid substitutions in the hemagglutinin of influenza A (H3N2) viruses on vaccine production and antibody response. *Vaccine* 28 (24): 4079–4085. doi:10.1016/j.vaccine.2010.03.078. (Cit. on pp. 14, 19).

Choppin, P. W., J. S. Murphy, and I. Tamm. 1960. Studies of two kinds of virus particles which comprise influenza A virus strains. III. Morphological and functional traits. *Journal of Experimental Medicine* 112 (5): 945–952. doi:10.1084/jem.112.5.945. (Cit. on p. 3).

Chu, C. M., I. M. Dawson, and W. J. Elford. 1949. Filamentous forms associated with newly isolated influenza virus. *The Lancet* 253 (6554): 602–603. doi:10.1016/S0140-6736(49)91699-2. (Cit. on p. 3).

Colman, P. M., and M. C. Lawrence. 2003. The structural biology of type I viral membrane fusion. *Nature reviews. Molecular cell biology* 4 (4): 309–19. doi:10.1038/nrm1076. (Cit. on p. 5).

Connor, R. J., et al. 1994. Receptor Specificity in Human, Avian, and Equine H2 and H3 Influenza Virus Isolates. *Virology* 205 (1): 17–23. doi:10.1006/VIRO.1994.1615. (Cit. on p. 16).

Coudeville, L., et al. 2010. Relationship between haemagglutination-inhibiting antibody titres and clinical protection against influenza: development and application of a bayesian random-effects model. *BMC medical research methodology* 10:18. doi:10.1186/1471-2288-10-18. (Cit. on pp. 10, 56).

Cox, M. M. J., et al. 2015. Safety, efficacy, and immunogenicity of Flublok in the prevention of seasonal influenza in adults. *Therapeutic Advances in Vaccines* 3 (4): 97. doi:10.1177/2051013615595595. (Cit. on p. 19).

Cox, R. J. 2013. Correlates of protection to influenza virus, where do we go from here? *Human vaccines & immunotherapeutics* 9 (2): 405–8. doi:10.4161/hv.22908. (Cit. on p. 10).

Dawood, F. S., et al. 2012. Estimated global mortality associated with the first 12 months of 2009 pandemic influenza A H1N1 virus circulation: A modelling study. *The Lancet Infectious Diseases* 12 (9): 687–695. doi:10.1016/S1473-3099(12)70121-4. (Cit. on p. 8).

De Serres, G., et al. 2013. The test-negative design: validity, accuracy and precision of vaccine efficacy estimates compared to the gold standard of randomised placebo-controlled clinical trials. *Eurosurveillance* 18 (37): 20585. doi:10.2807/1560-7917.ES2013.18.37.20585. (Cit. on p. 43).

Devlin, B., and K. Roeder. 1999. Genomic control for association studies. *Biometrics* 55 (4): 997–1004. (Cit. on p. 27).

Devore, J. L. 2000. *Probability and Statistics for Engineering and the Sciences*. 5th. Belmont, CA: Wadsworth. (Cit. on p. 25).

Donatelli, I., et al. 2016. Human–Animal Interface: The Case for Influenza Interspecies Transmission, 17–33. Springer, Cham. doi:10.1007/5584_2016_136. (Cit. on p. 2).

Drake, J. W. 1993. Rates of spontaneous mutation among RNA viruses. *Proceedings of the National Academy of Sciences of the United States of America* 90 (9): 4171–5. doi:10.1073/pnas.90.9.4171. (Cit. on p. 17).

Duffy, S., L. A. Shackelton, and E. C. Holmes. 2008. Rates of evolutionary change in viruses: patterns and determinants. *Nature reviews. Genetics* 9 (4): 267–76. doi:10.1038/nrg2323. (Cit. on p. 17).

Ekiert, D. C., et al. 2012. Cross-neutralization of influenza A viruses mediated by a single antibody loop. *Nature* 489 (7417): 526–532. doi:10.1038/nature11414. (Cit. on pp. 101, 103).

Erbelding, E. J., et al. 2018. A Universal Influenza Vaccine: The Strategic Plan for the National Institute of Allergy and Infectious Diseases. *The Journal of Infectious Diseases* 218 (3): 347–354. doi:10.1093/infdis/jiy103. (Cit. on p. 43).

Fauchère, J. L., et al. 1988. Amino acid side chain parameters for correlation studies in biology and pharmacology. *International journal of peptide and protein research* 32 (4): 269–78. (Cit. on p. 86).

Fielding, J. E., et al. 2016. Effectiveness of seasonal influenza vaccine in Australia, 2015: An epidemiological, antigenic and phylogenetic assessment. *Vaccine* 34 (41): 4905–4912. doi:10.1016/j.vaccine.2016.08.067. (Cit. on pp. 43, 45).

Fields, B. N., D. M. Knipe, and P. M. Howley. 2013. *Fields virology*. 82. Wolters Kluwer Health/Lippincott Williams & Wilkins. (Cit. on pp. 3, 5).

Flannery, B., et al. 2016. Enhanced Genetic Characterization of Influenza A(H3N2) Viruses and Vaccine Effectiveness by Genetic Group, 2014–2015. *Journal of Infectious Diseases* 214 (7): 1010–1019. doi:10.1093/infdis/jiw181. (Cit. on p. 43).

Flannery, B., et al. 2017. Interim Estimates of 2017–18 Seasonal Influenza Vaccine Effectiveness — United States, February 2018. *Morbidity and Mortality Weekly Report* 67 (6): 180–185. (Cit. on p. 51).

Fonville, J. M., et al. 2014. Antibody landscapes after influenza virus infection or vaccination. *Science* 346 (6212): 996–1000. doi:10.1126/science.1256427. (Cit. on pp. 14, 82, 107).

Fonville, J. M., et al. 2016. Antigenic maps of influenza A(H3N2) produced with human antisera obtained after primary infection. *Journal of Infectious Diseases* 213 (1): 31–38. doi:10.1093/infdis/jiv367. (Cit. on pp. 14, 16).

Foppa, I. M., et al. 2013. The case test-negative design for studies of the effectiveness of influenza vaccine. *Vaccine* 31 (30): 3104–3109. doi:10.1016/j.vaccine.2013.04.026. (Cit. on p. 43).

Fouchier, R. A. M., and Y. Guan. 2013. Ecology and evolution of influenza viruses in wild and domestic birds. In *Textbook of Influenza*, 173–189. Oxford, UK: John Wiley & Sons, Ltd. doi:10.1002/9781118636817.ch11. (Cit. on p. 2).

Francis, T., et al. 1945. Protective Effect of Vaccination Against Induced Influenza A. *The Journal of clinical investigation* 24 (4): 536–46. doi:10.1172/JCI101633. (Cit. on p. 14).

Frost, W. 1920. Statistics of influenza morbidity with special reference to certain factors in case incidence and case fatality. *Public Health Rep.* 35:584–597. (Cit. on p. 7).

Fukushima, W., and Y. Hirota. 2017. Basic principles of test-negative design in evaluating influenza vaccine effectiveness. *Vaccine* 35 (36): 4796–4800. doi:10.1016/J.VACCINE.2017.07.003. (Cit. on p. 46).

Galwey, N. W. 2009. A new measure of the effective number of tests, a practical tool for comparing families of non-independent significance tests. *Genetic Epidemiology* 33 (7): 559–568. doi:10.1002/gepi.20408. (Cit. on p. 33).

Garten, R. J., et al. 2009. Antigenic and genetic characteristics of swine-origin 2009 A(H1N1) influenza viruses circulating in humans. *Science (New York, N.Y.)* 325 (5937): 197–201. doi:10.1126/science.1176225. (Cit. on p. 14).

Gherasim, A., et al. 2017. Effect of previous and current vaccination against influenza A(H1N1)pdm09, A(H3N2), and B during the post-pandemic period 2010-2016 in Spain. *PLoS ONE* 12 (6): 1–16. doi:10.1371/journal.pone.0179160. (Cit. on p. 51).

Gulati, S., et al. 2013. Human H3N2 Influenza Viruses Isolated from 1968 To 2012 Show Varying Preference for Receptor Substructures with No Apparent Consequences for Disease or Spread. *PloS one* 8 (6): e66325. doi:10.1371/journal.pone.0066325. (Cit. on p. 15).

Hirst, G. K. 1941. The Agglutination of Red Cells by Allantoic Fluid of Chick Embryos Infected with Influenza Virus. *Science (New York, N.Y.)* 94 (2427): 22–3. doi:10.1126/science.94.2427.22. (Cit. on p. 10).

– . 1942. The Quantitative Determination of Influenza Virus and Antibodies by Means of Red Cell Agglutination. *The Journal of Experimental Medicine* 75 (1): 49–64. doi:10.1084/jem.75.1.49. (Cit. on p. 10).

Hobson, D., et al. 1972. The role of serum haemagglutination-inhibiting antibody in protection against challenge infection with influenza A2 and B viruses. *The Journal of Hygiene* 70 (4): 767–77. (Cit. on p. 10).

Hu, J. K., X. Wang, and P. Wang. 2014. Testing gene-gene interactions in genome wide association studies. *Genetic epidemiology* 38 (2): 123–34. doi:10.1002/gepi.21786. (Cit. on p. 34).

Iuliano, A. D., et al. 2018. Estimates of global seasonal influenza-associated respiratory mortality: a modelling study. *The Lancet* 391 (10127): 1285–1300. doi:10.1016/S0140-6736(17)33293-2. (Cit. on p. 7).

Jackson, M. L., and J. C. Nelson. 2013. The test-negative design for estimating influenza vaccine effectiveness. *Vaccine* 31 (17): 2165–2168. doi:10.1016/j.vaccine.2013.02.053. (Cit. on p. 43).

Jackson, M. L., et al. 2017. Influenza Vaccine Effectiveness in the United States during the 2015–2016 Season. *New England Journal of Medicine* 377 (6): 534–543. doi:10.1056/NEJMoa1700153. (Cit. on p. 51).

Jia, N., et al. 2014. Glycomic characterization of respiratory tract tissues of ferrets: implications for its use in influenza virus infection studies. *The Journal of Biological Chemistry* 289 (41): 28489. doi:10.1074/JBC.M114.588541. (Cit. on p. 16).

Jones, E., T. Oliphant, P. Peterson, et al. 2001. *SciPy: Open source scientific tools for Python*. (Cit. on p. 89).

Jong, J. C. de, et al. 2007. Antigenic and genetic evolution of swine influenza A (H3N2) viruses in Europe. *Journal of virology* 81 (8): 4315–22. doi:10.1128/JVI.02458-06. (Cit. on pp. 13, 16).

Jong, J. de, et al. 2000. Mismatch between the 1997/1998 influenza vaccine and the major epidemic A (H3N2) virus strain as the cause of an inadequate vaccine-induced antibody response to this strain in the elderly. *Journal of Medical Virology*: 94–99. (Cit. on p. 45).

Kanegae, Y., et al. 1990. Evolutionary pattern of the hemagglutinin gene of influenza B viruses isolated in Japan: cocirculating lineages in the same epidemic season. *Journal of Virology* 64 (6): 2860. (Cit. on p. 7).

Kang, H. M., et al. 2010. Variance component model to account for sample structure in genome-wide association studies. *Nature Publishing Group* 42 (4): 348–354. doi:10.1038/ng.548. (Cit. on pp. 27, 28).

Katoh, K., and D. M. Standley. 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution* 30 (4): 772–780. doi:10.1093/molbev/mst010. (Cit. on p. 49).

Katzelnick, L. C., et al. 2015. Dengue viruses cluster antigenically but not as discrete serotypes. *Science (New York, N.Y.)* 349 (6254): 1338–1343. doi:10.1126/science.aac5017. (Cit. on pp. 14, 29).

Kawakami, C., et al. 2019. Genetic and antigenic characterisation of influenza A(H3N2) viruses isolated in Yokohama during the 2016/17 and 2017/18 influenza seasons. *Eurosurveillance* 24 (6). doi:10.2807/1560-7917.ES.2019.24.6.1800467. (Cit. on p. 14).

Klimov, A. I., et al. 2012. WHO recommendations for the viruses to be used in the 2012 Southern Hemisphere Influenza Vaccine: Epidemiology, antigenic and genetic characteristics of influenza A(H1N1)pdm09, A(H3N2) and B influenza viruses collected from February to September 2011. *Vaccine* 30 (45): 6461–6471. doi:10.1016/j.vaccine.2012.07.089. (Cit. on p. 73).

Koel, B. F., et al. 2013. Substitutions Near the Receptor Binding Site Determine Major Antigenic Change During Influenza Virus Evolution. *Science* 342 (6161): 976–979. doi:10.1126/science.1244730. (Cit. on pp. 18, 31, 34, 37, 41, 60, 63–67, 78, 81, 84, 85, 88–90, 101, 108, 112).

Koel, B. F., et al. 2014. Antigenic variation of clade 2.1 H5N1 virus is determined by a few amino acid substitutions immediately adjacent to the receptor binding site. *mBio* 5 (3). doi:10.1128/mBio.01070-14. (Cit. on pp. 14, 16, 17, 34, 41, 64–66, 78).

Koelle, K., et al. 2006. Epochal evolution shapes the phylodynamics of interpandemic influenza A (H3N2) in humans. *Science (New York, N.Y.)* 314 (5807): 1898–903. doi:10.1126/science.1132745. (Cit. on p. 17).

Korte, A., et al. 2012. A mixed-model approach for genome-wide association studies of correlated traits in structured populations. *Nature genetics* 44 (9): 1066–71. doi:10.1038/ng.2376. (Cit. on pp. 27, 29).

Krammer, F., et al. 2018. Influenza. *Nature Reviews. Disease Primers* 4 (3): 1–21. (Cit. on pp. 1, 7).

Krammer, F. 2019. The human antibody response to influenza A virus infection and vaccination. *Nature Reviews Immunology*: 1. doi:10.1038/s41577-019-0143-6. (Cit. on pp. 14, 19, 43).

Krammer, F., and P. Palese. 2015. Advances in the development of influenza virus vaccines. *Nature Reviews Drug Discovery* 14 (3): 167–182. doi:10.1038/nrd4529. (Cit. on p. 18).

Kruschke, J. K. 2014. *Doing Bayesian data analysis: a tutorial with R, JAGS, and Stan*. 2nd ed. 759. Elsevier. (Cit. on p. 59).

Kyte, J., and R. F. Doolittle. 1982. A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology* 157 (1): 105–132. doi:10.1016/0022-2836(82)90515-0. (Cit. on p. 86).

Larsson, A. 2014. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* 30 (22): 3276–3278. doi:10.1093/bioinformatics/btu531. (Cit. on p. 49).

Lehner, B. 2011. Molecular mechanisms of epistasis within and between genes. *Trends in Genetics* 27 (8): 323–31. doi:10.1016/j.tig.2011.05.007. (Cit. on p. 34).

Lewis, N. S., et al. 2011. Antigenic and genetic evolution of equine influenza A (H3N8) virus from 1968 to 2007. *Journal of virology* 85 (23): 12742–9. doi:10.1128/JVI.05319-11. (Cit. on pp. 14, 29, 34, 41, 65, 66, 78).

Lewis, N. S., et al. 2014. Substitutions near the hemagglutinin receptor-binding site determine the antigenic evolution of influenza A H3N2 viruses in U.S. swine. *Journal of virology* 88 (9): 4752–63. doi:10.1128/JVI.03805-13. (Cit. on pp. 13, 29, 41, 64, 65, 78).

Li, C., et al. 2016. Selection of antigenically advanced variants of seasonal influenza viruses. *Nature Microbiology* 1 (6): 16058. doi:10.1038/nmicrobiol.2016.58. (Cit. on pp. 36, 37, 40, 41, 66, 67, 72, 73, 117–121).

Lin, Y. P., et al. 2012. Evolution of the receptor binding properties of the influenza A(H3N2) hemagglutinin. *Proceedings of the National Academy of Sciences of the United States of America* 109 (52): 21474–9. doi:10.1073/pnas.1218841110. (Cit. on p. 15).

Lin, Y. P., et al. 2010. Neuraminidase receptor binding variants of human influenza A(H3N2) viruses resulting from substitution of aspartic acid 151 in the catalytic site: a role in virus attachment? *Journal of virology* 84 (13): 6769–81. doi:10.1128/JVI.00458-10. (Cit. on p. 4).

Linderman, S. L., et al. 2014. Potential antigenic explanation for atypical H1N1 infections among middle-aged adults during the 2013-2014 influenza season. *Proceedings of the National Academy of Sciences of the United States of America* 111 (44): 15798–803. doi:10.1073/pnas.1409171111. (Cit. on p. 16).

Lindstrom, S. E., N. J. Cox, and A. Klimov. 2004. Genetic analysis of human H2N2 and early H3N2 influenza viruses, 1957–1972: evidence for genetic divergence and multiple reassortment events. *Virology* 328 (1): 101–119. doi:10.1016/J.VIROL.2004.06.009. (Cit. on p. 3).

Linster, M., et al. 2019. The Molecular Basis for Antigenic Drift of Human A/H2N2 Influenza Viruses. Ed. by A. García-Sastre. *Journal of Virology* 93 (8). doi:10.1128/JVI.01907-18. (Cit. on p. 14).

Lippert, C., et al. 2011. FaST linear mixed models for genome-wide association studies. *Nature Methods* 8 (10): 833–835. doi:10.1038/nmeth.1681. (Cit. on pp. 27, 28).

Lippert, C., et al. 2014. LIMIX : genetic analysis of multiple traits. *bioRxiv*: 1–26. doi:https://doi.org/10.1101/003905. (Cit. on p. 29).

Liu, D. C., and J. Nocedal. 1989. On the limited memory BFGS method for large scale optimization. *Mathematical Programming* 45 (1-3): 503–528. doi:10.1007/BF01589116. (Cit. on p. 12).

Lyons, D., et al. 2018. Mutation and Epistasis in Influenza Virus Evolution. *Viruses* 10 (8): 407. doi:10.3390/v10080407. (Cit. on p. 43).

Magill, T. P., and T. Francis. 1936. Antigenic Differences in Strains of Human Influenza Virus. *Experimental Biology and Medicine* 35 (3): 463–466. doi:10.3181/00379727-35-9018C. (Cit. on p. 8).

Marchini, J., P. Donnelly, and L. R. Cardon. 2005. Genome-wide strategies for detecting multiple loci that influence complex diseases. *Nature Genetics* 37 (4): 413–417. doi:10.1038/ng1537. (Cit. on p. 34).

Matsuzaki, Y., et al. 2006. Clinical Features of Influenza C Virus Infection in Children. *The Journal of Infectious Diseases* 193 (9): 1229–1235. doi:10.1086/502973. (Cit. on p. 7).

McCullers, J. A., T. Saito, and A. R. Iverson. 2004. Multiple genotypes of influenza B virus circulated between 1979 and 2003. *Journal of virology* 78 (23): 12817–28. doi:10.1128/JVI.78.23.12817-12828.2004. (Cit. on p. 7).

Meyer, H. V. 2017. "Genetic association of high dimensional traits". PhD thesis, University of Cambridge. (Cit. on p. 29).

Mögling, R., et al. 2017. Neuraminidase-mediated haemagglutination of recent human influenza A(H3N2) viruses is determined by arginine 150 flanking the neuraminidase catalytic site. *J.Gen.Virol*: 1–8. doi:10.1099/jgv.0.000809. (Cit. on pp. 4, 118).

Molinari, N.-A. M., et al. 2007. The annual impact of seasonal influenza in the US: measuring disease burden and costs. *Vaccine* 25 (27): 5086–96. doi:10.1016/j.vaccine.2007.03.046. (Cit. on p. 8).

Morris, J. A., and M. J. Gardner. 1988. Calculating confidence intervals for relative risks (odds ratios) and standardised ratios and rates. *BMJ (Clinical research ed.)* 296 (6632): 1313–1316. doi:10.1136/bmj.296.6632.1313. (Cit. on p. 47).

Mosterín, A. 2014. [A systematic approach to improving influenza vaccine efficacy]. *Unpublished raw data*. (Cit. on p. 8).

Murphy, K. P. 2012. *Machine Learning: A Probabilistic Perspective*. The MIT Press. (Cit. on p. 91).

Nachman, M. W., and S. L. Crowell. 2000. Estimate of the Mutation Rate per Nucleotide in Humans. *Genetics* 156:297–304. (Cit. on p. 17).

Nair, H., et al. 2011. Global burden of respiratory infections due to seasonal influenza in young children: A systematic review and meta-analysis. *The Lancet* 378 (9807): 1917–1930. doi:10.1016/S0140-6736(11)61051-9. (Cit. on p. 8).

Nelson, M. I., and E. C. Holmes. 2007. The evolution of epidemic influenza. *Nature Reviews Genetics* 8 (3): 196–205. doi:10.1038/nrg2053. (Cit. on p. 16).

Neumann, G., and Y. Kawaoka. 2015. Transmission of influenza A viruses. *Virology* 479-480:234–246. doi:10.1016/j.virol.2015.03.009. (Cit. on p. 7).

Nfon, C. K., et al. 2011. Characterization of H1N1 swine influenza viruses circulating in Canadian pigs in 2009. *Journal of virology* 85 (17): 8667–79. doi:10.1128/JVI.00801-11. (Cit. on p. 13).

Oh, D. Y., and A. C. Hurt. 2016. Using the Ferret as an Animal Model for Investigating Influenza Antiviral Effectiveness. *Frontiers in microbiology* 7:80. doi:10.3389/fmicb.2016.00080. (Cit. on p. 16).

Paules, C. I., et al. 2017. Chasing Seasonal Influenza — The Need for a Universal Influenza Vaccine. *New England Journal of Medicine*. doi:10.1056/NEJMp1714916. (Cit. on p. 43).

Payne, A. M. 1953. The influenza programme of WHO. *Bulletin of the World Health Organization* 8 (5-6): 755–74. (Cit. on pp. 9, 20).

Pedregosa, F., et al. 2011. Scikit-learn: Machine Learning in {P}ython. *Journal of Machine Learning Research* 12:2825–2830. (Cit. on p. 91).

Philippe, D., et al. 2009. Global immunization: status, progress, challenges and future. *BMC International Health and Human Rights* 9 (Suppl 1): S2. doi:10.1186/1472-698x-9-s1-s2. (Cit. on p. 9).

Pinheiro, J. C., and D. M. Bates. 2000. *Mixed-Effects Models in S and S-PLUS*. Statistics and Computing. New York: Springer-Verlag. doi:10.1007/b98882. (Cit. on p. 25).

Price, A. L., et al. 2006. Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics* 38 (8): 904–909. doi:10.1038/ng1847. (Cit. on p. 27).

Pritchard, J. K., et al. 2000. Association Mapping in Structured Populations. *The American Journal of Human Genetics* 67 (1): 170–181. doi:10.1086/302959. (Cit. on p. 27).

Pybus, O. G., A. J. Tatem, and P. Lemey. 2015. Virus evolution and transmission in an ever more connected world. *Proceedings of the Royal Society B: Biological Sciences* 282 (1821): 1–10. doi:10.1098/rspb.2014.2878. (Cit. on p. 20).

Ray, T., et al. 2019a. Depletion-of-susceptibles Bias in Analyses of Intra-season Waning of Influenza Vaccine Effectiveness. *Clinical Infectious Diseases*: 1–3. doi:10.1093/cid/ciz706. (Cit. on p. 61).

Ray, T., et al. 2019b. Intraseason waning of influenza vaccine effectiveness. *Clinical Infectious Diseases* 68 (10): 1623–1630. doi:10.1093/cid/ciy770. (Cit. on p. 61).

Robinson, G. K. 1991. That BLUP is a good thing: the estimation of random effects. *Statistical Science* 6 (1): 15–32. (Cit. on p. 35).

Roseman, M. A. 1988. Hydrophilicity of polar amino acid side-chains is markedly reduced by flanking peptide bonds. *Journal of Molecular Biology* 200 (3): 513–522. doi:10.1016/0022-2836(88)90540-2. (Cit. on p. 86).

Roser, M. 2019. Life Expectancy. *Our World in Data*. (Cit. on p. 9).

Roser, M., H. Ritchie, and E. Ortiz-Ospina. 2019. World Population Growth. *Our World in Data*. (Cit. on p. 8).

Rota, P. A., et al. 1992. Antigenic and Genetic Characterization of the Haemagglutinins of Recent Cocirculating Strains of Influenza B Virus. *Journal of General Virology* 73 (10): 2737–2742. doi:10.1099/0022-1317-73-10-2737. (Cit. on p. 7).

Russell, C. A., et al. 2008a. Influenza vaccine strain selection and recent studies on the global migration of seasonal influenza viruses. *Vaccine* 26:D31–D34. doi:10.1016/j.vaccine.2008.07.078. (Cit. on pp. 14, 20, 47).

– . 2008b. The global circulation of seasonal influenza A (H3N2) viruses. *Science (New York, N.Y.)* 320 (5874): 340–6. doi:10.1126/science.1154137. (Cit. on pp. 14, 20, 30, 36, 37, 39, 66, 67, 69, 72, 117–121).

Russell, C. A., et al. 2012. The potential for respiratory droplet-transmissible A/H5N1 influenza virus to evolve in a mammalian host. *Science (New York, N.Y.)* 336 (6088): 1541–7. doi:10.1126/science.1222526. (Cit. on pp. 17, 18).

Salk, J. E., and P. C. Suriano. 1949. Importance of antigenic composition of influenza virus vaccine in protecting against the natural disease; observations during the winter of 1947-1948. *American Journal of Public Health and the Nation's Health* 39 (3): 345–55. doi:10.2105/ajph.39.3.345. (Cit. on p. 9).

Salmon, D. A., et al. 1999. Health Consequences of Religious and Philosophical Exemptions From Immunization Laws. *JAMA* 282 (1): 47. doi:10.1001/jama.282.1.47. (Cit. on p. 9).

Salmon, D. A., et al. 2015. Vaccine hesitancy: Causes, consequences, and a call to action. *Vaccine* 33:D66–D71. doi:10.1016/J.VACCINE.2015.09.035. (Cit. on p. 9).

Salvatier, J., T. V. Wiecki, and C. Fonnesbeck. 2016. Probabilistic programming in Python using PyMC3. *PeerJ Computer Science* 2:e55. doi:10.7717/peerj-cs.55. (Cit. on p. 57).

Sauter, N. K., et al. 1992. Binding of Influenza Virus Hemagglutinin to Analogs of Its Cell-Surface Receptor, Sialic Acid: Analysis by Proton Nuclear Magnetic Resonance Spectroscopy and X-ray Crystallography. *Biochemistry* 31 (40): 9609–9621. doi:10.1021/bi00155a013. (Cit. on p. 5).

Sauter, N. K., et al. 1989. Hemagglutinins from Two Influenza Virus Variants Bind to Sialic Acid Derivatives with Millimolar Dissociation Constants: A 500-MHz Proton Nuclear Magnetic Resonance Study. *Biochemistry* 28 (21): 8388–8396. doi:10.1021/bi00447a018. (Cit. on p. 5).

Schrauwen, E. J. A., et al. 2014. Determinants of virulence of influenza A virus. *European Journal of Clinical Microbiology and Infectious Diseases* 33 (4): 479–490. doi:10.1007/s10096-013-1984-8. (Cit. on p. 7).

Schrödinger LLC. 2015. The PyMOL Molecular Graphics System. (Cit. on pp. 80, 103).

Scott, D. 1992. *Multivariate Density Estimation: Theory, Practice, and Visualization*. John Wiley & Sons, New York, Chichester. (Cit. on pp. 89, 91).

Shu, Y., and J. McCauley. 2017. GISAID: Global initiative on sharing all influenza data – from vision to reality. *Eurosurveillance* 22 (13). doi:10.2807/1560-7917.ES.2017.22.13.30494. (Cit. on p. 47).

Skehel, J. J., and D. C. Wiley. 2000. Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. *Annual review of biochemistry* 69:531–69. doi:10.1146/annurev.biochem.69.1.531. (Cit. on pp. 4, 5).

Skowronski, D. M., et al. 2007. Estimating vaccine effectiveness against laboratory-confirmed influenza using a sentinel physician network: Results from the 2005–2006 season of dual A and B vaccine mismatch in Canada. *Vaccine* 25 (15): 2842–2851. doi:10.1016/j.vaccine.2006.10.002. (Cit. on pp. 43, 45).

Skowronski, D. M., et al. 2017. Beyond Antigenic Match: Possible Agent-Host and Immuno-epidemiological Influences on Influenza Vaccine Effectiveness During the 2015–2016 Season in Canada. *The Journal of Infectious Diseases* 216 (12): 1487–1500. doi:10.1093/infdis/jix526. (Cit. on p. 45).

Skowronski, D. M., et al. 2014. Low 2012-13 influenza vaccine effectiveness associated with mutation in the egg-adapted H3N2 vaccine strain not antigenic drift in circulating viruses. *PloS one* 9 (3): e92153. doi:10.1371/journal.pone.0092153. (Cit. on p. 60).

Smith, D., et al. 2004. Mapping the Antigenic and Genetic Evolution of Influenza Virus. *Science* 305 (5682): 371–376. doi:10.1126/science.1097211. (Cit. on pp. 11–14, 16, 29, 36–38, 41, 66, 111, 117–119, 121).

Smith, G. J. D., et al. 2009. Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. *Nature* 459 (7250): 1122–1125. doi:10.1038/nature08182. (Cit. on p. 3).

Smith, W., C. H. Andrewes, and P. P. Laidlaw. 1933. A virus obtained from influenza patients. *Lancet* 222:66–68. (Cit. on pp. 8, 15).

Smith, W., and C. H. Andrewes. 1938. Serological races of influenza virus. *British Journal of Experimental Pathology* 19 (5): 293–314. (Cit. on p. 9).

Spielman, R. S., R. E. McGinnis, and W. J. Ewens. 1993. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *American journal of human genetics* 52 (3): 506–16. (Cit. on p. 27).

SPSN. 2017. *Canadian Sentinel Practitioner Surveillance Network (SPSN) influenza vaccine effectiveness estimates % (95% CI), 2004-05 to 2016-17 seasons*. doi:10.1136/bmj.c7297. (Cit. on p. 51).

Sullivan, S. G., E. J. Tchetgen Tchetgen, and B. J. Cowling. 2016. Theoretical Basis of the Test-Negative Study Design for Assessment of Influenza Vaccine Effectiveness. *American journal of epidemiology* 184 (5): 345–53. doi:10.1093/aje/kww064. (Cit. on p. 43).

Taubenberger, J. K., et al. 2001. Integrating historical, clinical and molecular genetic data in order to explain the origin and virulence of the 1918 Spanish influenza virus. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 356 (1416): 1829–39. doi:10.1098/rstb.2001.1020. (Cit. on p. 16).

Taubenberger, J. K., and D. M. Morens. 2006. 1918 Influenza: the mother of all pandemics. *Emerging infectious diseases* 12 (1): 15–22. doi:10.3201/eid1201.050979. (Cit. on p. 8).

Thi Nguyen, D., et al. 2018. Antigenic characterization of highly pathogenic avian influenza A(H5N1) viruses with chicken and ferret antisera reveals clade-dependent variation in hemagglutination inhibition profiles. *Emerging Microbes & Infections* 7 (1): 1–15. doi:10.1038/s41426-018-0100-7. (Cit. on p. 14).

Tong, S., et al. 2012. A distinct lineage of influenza A virus from bats. *Proceedings of the National Academy of Sciences of the United States of America* 109 (11): 4269–74. doi:10.1073/pnas.1116200109. (Cit. on p. 2).

Tong, S., et al. 2013. New world bats harbor diverse influenza A viruses. *PLoS pathogens* 9 (10): e1003657. doi:10.1371/journal.ppat.1003657. (Cit. on p. 2).

Tricco, A. C., et al. 2013. Comparing influenza vaccine efficacy against mismatched and matched strains: a systematic review and meta-analysis. *BMC Medicine* 11 (1): 153. doi:10.1186/1741-7015-11-153. (Cit. on pp. 44, 45).

Viboud, C., et al. 2016. Global Mortality Impact of the 1957–1959 Influenza Pandemic. *Journal of Infectious Diseases* 213 (5): 738–745. doi:10.1093/infdis/jiv534. (Cit. on p. 8).

Viboud, C., et al. 2005. Multinational Impact of the 1968 Hong Kong Influenza Pandemic: Evidence for a Smoldering Pandemic. *The Journal of Infectious Diseases* 192 (2): 233–248. doi:10.1086/431150. (Cit. on p. 8).

Webster, R. G., et al. 1992. Evolution and ecology of influenza A viruses. *Microbiological reviews* 56 (1): 152–79. (Cit. on pp. 2, 7).

Webster, R. G., et al. 2013. *Textbook of influenza*. 2nd ed. Ed. by R. G. Webster et al. 522. Wiley. (Cit. on p. 3).

White, M. C., et al. 2019. H5N8 and H7N9 packaging signals constrain HA reassortment with a seasonal H3N2 influenza A virus. *Proceedings of the National Academy of Sciences* 116 (10): 4611–4618. doi:10.1073/pnas.1818494116. (Cit. on p. 6).

Widjaja, L., et al. 2006. Molecular changes associated with adaptation of human influenza A virus in embryonated chicken eggs. *Virology* 350 (1): 137–145. doi:10.1016/J.VIROL.2006.02.020. (Cit. on p. 14).

Wiley, D. C., I. A. Wilson, and J. J. Skehel. 1981. Structural identification of the antibody-binding sites of Hong Kong influenza haemagglutinin and their involvement in antigenic variation. *Nature* 289 (5796): 373–378. doi:10.1038/289373a0. (Cit. on pp. 4, 31, 41, 83).

Wilson, I. A., and N. J. Cox. 1990. Structural basis of immune recognition of influenza virus hemagglutinin. *Annual review of immunology* 8:737–71. doi:10.1146/annurev.iy.08.040190.003513. (Cit. on pp. 31, 41, 83).

Wood, J. 1998. Standardization of inactivated influenza vaccine. In *Textbook of Influenza*, ed. by K. Nicholson, R. Webster, and A. Hay, 333–345. London, UK: Blackwell Science Ltd. (Cit. on p. 19).

World Health Organisation. 2014. *Influenza (Seasonal) Fact Sheet*. Visited on 11/08/2014. (Cit. on p. 7).

World Health Organization. 1980. A revision of the system of nomenclature for influenza viruses: a WHO Memorandum*. *Bulletin of the World Health Organization* 58 (4): 585–591. (Cit. on p. 2).

Worobey, M., G.-Z. Han, and A. Rambaut. 2014. Genesis and pathogenesis of the 1918 pandemic H1N1 influenza A virus. *Proceedings of the National Academy of Sciences of the United States of America* 111 (22): 8107–12. doi:10.1073/pnas.1324197111. (Cit. on p. 3).

Wu, J. T., et al. 2010. The Infection Attack Rate and Severity of 2009 Pandemic H1N1 Influenza in Hong Kong. *Clinical Infectious Diseases* 51 (10): 1184–1191. doi:10.1086/656740. (Cit. on p. 7).

Xie, H., et al. 2015. H3N2 Mismatch of 2014–15 Northern Hemisphere Influenza Vaccines and Head-to-head Comparison between Human and Ferret Antisera derived Antigenic Maps. *Scientific Reports* 5 (1): 15279. doi:10.1038/srep15279. (Cit. on p. 45).

Xu, R., and I. A. Wilson. 2011. Structural characterization of an early fusion intermediate of influenza virus hemagglutinin. *Journal of virology* 85 (10): 5172–82. doi:10.1128/JVI.02430-10. (Cit. on p. 5).

Zhang, Z., et al. 2010. Mixed linear model approach adapted for genome-wide association studies. *Nature Genetics* 42 (4): 355–360. doi:10.1038/ng.546. (Cit. on pp. 27, 28).

Zhou, X., and M. Stephens. 2012. Genome-wide efficient mixed-model analysis for association studies. *Nature genetics* 44 (7): 821–4. doi:10.1038/ng.2310. (Cit. on pp. 27, 28).

Zimmerman, R. K., et al. 2016. 2014–2015 Influenza Vaccine Effectiveness in the United States by Vaccine Type. *Clinical Infectious Diseases* 63 (12): 1564–1573. doi:10.1093/cid/ciw635. (Cit. on pp. 43, 51).

Zost, S. J., et al. 2017. Contemporary H3N2 influenza viruses have a glycosylation site that alters binding of antibodies elicited by egg-adapted vaccine strains. *Proceedings of the National Academy of Sciences of the United States of America* 114 (47): 12578–12583. doi:10.1073/pnas.1712377114. (Cit. on p. 60).

COLOPHON

This document was typeset using the typographical look-and-feel `classicthesis` developed by André Miede and Ivo Pletikosić. The style was inspired by Robert Bringhurst's seminal book on typography "*The Elements of Typographic Style*". `classicthesis` is available for both LaTeX and LyX:

`https://bitbucket.org/amiede/classicthesis/`

Happy users of `classicthesis` usually send a real postcard to the author, a collection of postcards received so far is featured here:

`http://postcards.miede.de/`

Thank you very much for your feedback and contribution.

Final Version as of February 11, 2020 (classicthesis v4.6).